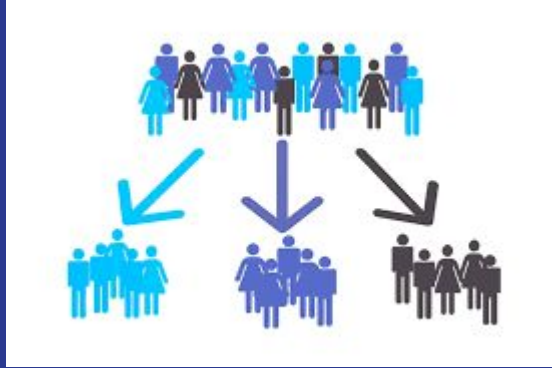


Segmentez des clients d'un site de e-commerce



Objectif : Segmentation et
Clustering des clients

Sommaire



- ❖ Problématique
- ❖ Analyser les données
- ❖ Description des données
- ❖ Analyse exploratoire des données
- ❖ Segmentation
- ❖ Clustering
- ❖ PEP8
- ❖ Conclusion et Suite du Projet



Problématique : Contexte

Objectif :

- ❖ **Comprendre** les différents types d'utilisateurs (Comportements, données personnelles)
- ❖ **Fournir** une description actionnable de la **segmentation** pour une optimisation optimale
- ❖ Proposer un **contrat de maintenance** basé sur une analyse de la **stabilité des segments** au cours du temps



Problématique : Mission

A partir des données récoltées :

Fournir aux équipes marketing de l'entreprise Olist une **Segmentation** des clients utilisables dans leurs campagnes de communication.



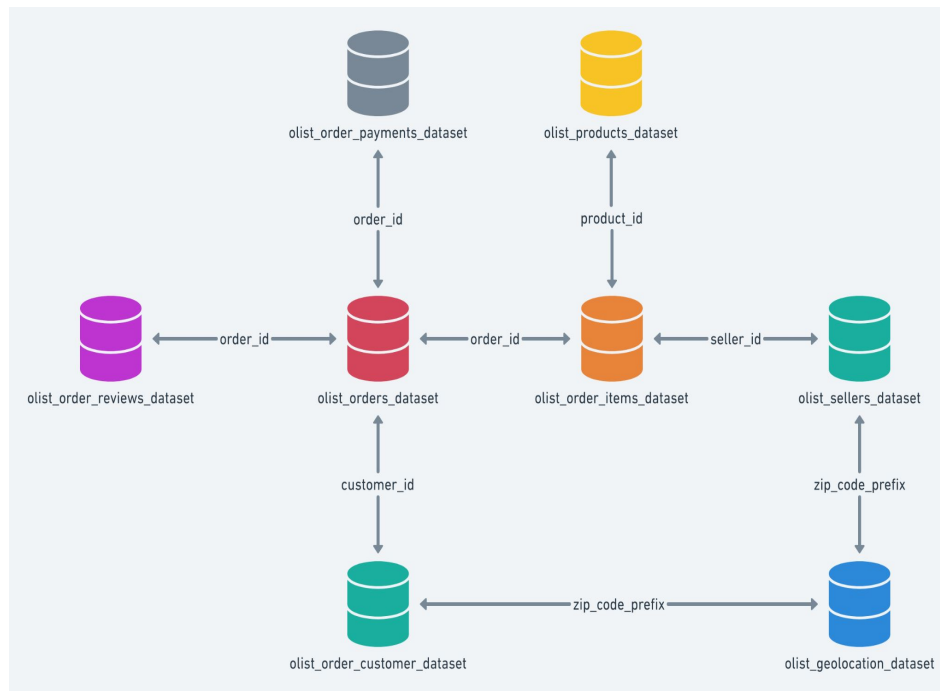
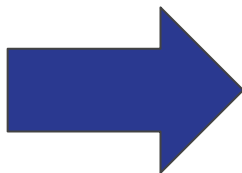
Analyser les données

- ❖ Réaliser une courte analyse exploratoire.
- ❖ Utiliser des méthodes non supervisés pour regrouper des clients de profils similaires. Ces catégories pourront être utilisées par l'équipe marketing pour mieux communiquer.
- ❖ Différencier les bons et moins clients en termes de commandes et de satisfaction.
- ❖ Mettre en place une recommandation de fréquence à laquelle la segmentation doit être mise à jour pour rester pertinente, afin de pouvoir effectuer un devis de contrat de maintenance.
- ❖ Respecter la convention PEP8 pour le code.



Descriptions des données

- ❖ **OLIST** : plateforme vitrine de e-commerce au Brésil (2016).
- ❖ Met en lien acheteurs et vendeurs . Commande, paiement, suivi de livraison. Notation, avis sur la commande....
- ❖ Les données sont disponibles à l'adresse suivante : [Olist Data](#)



INTERPRÉTATION

- Connaissance client améliorée.
- Segmentation intelligente.
- Analyse descriptive des profils.

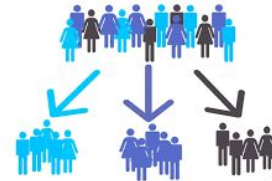
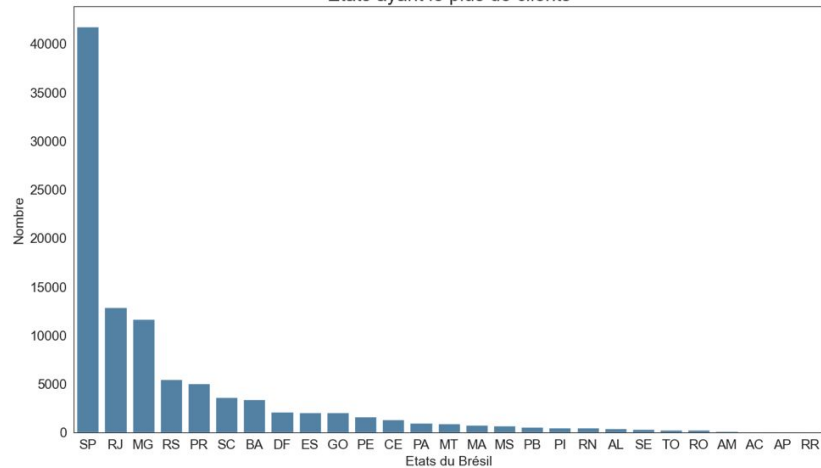


Diagramme de la matrice de la relation client :

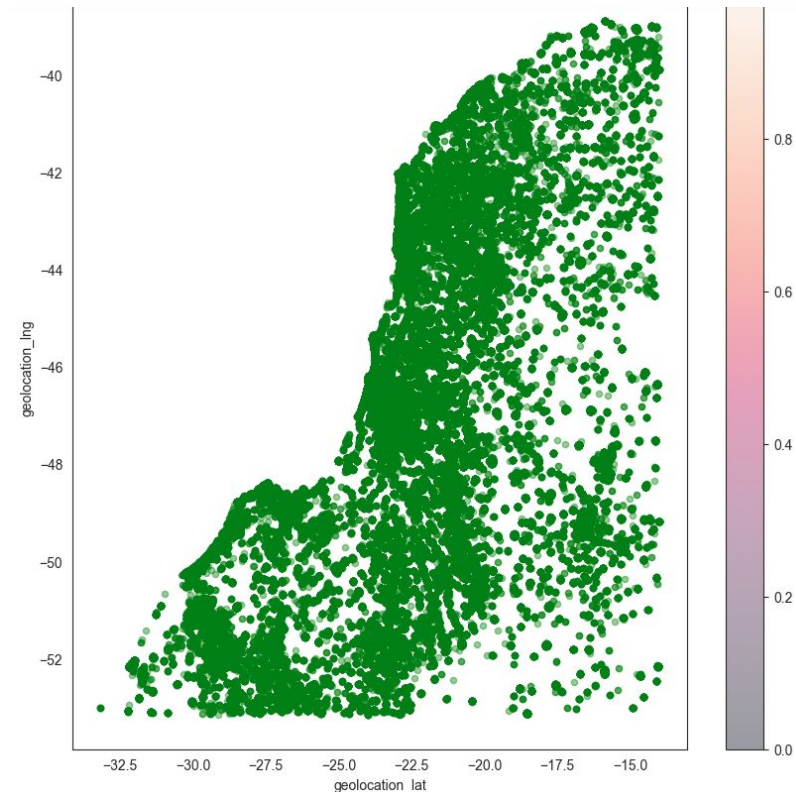
	Fréquence	Fréquence	Fréquence
Montant	Loyaux	Loyalistes potentiels	Perdus
	Champions	A réactiver	A risque
	Fréquence	Fréquence	Fréquence

Analyse exploratoire des données

Etats ayant le plus de clients



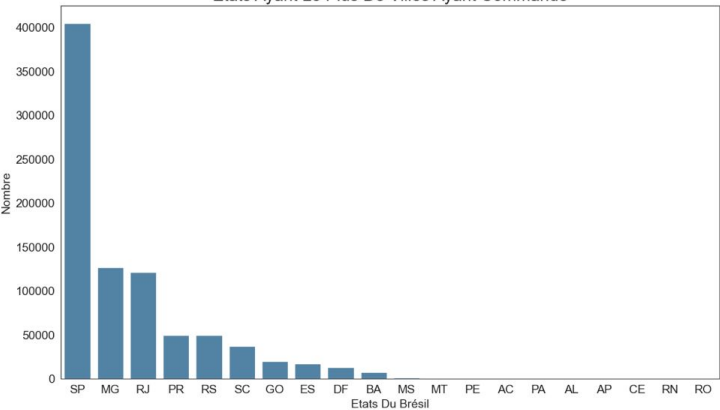
Localisation des clients



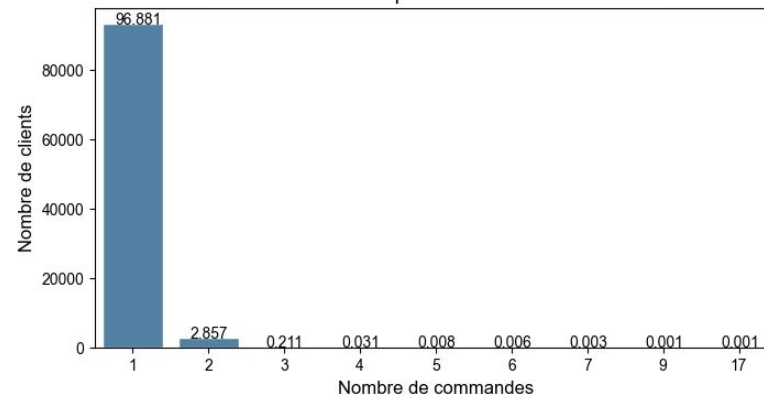
Analyse exploratoire des données



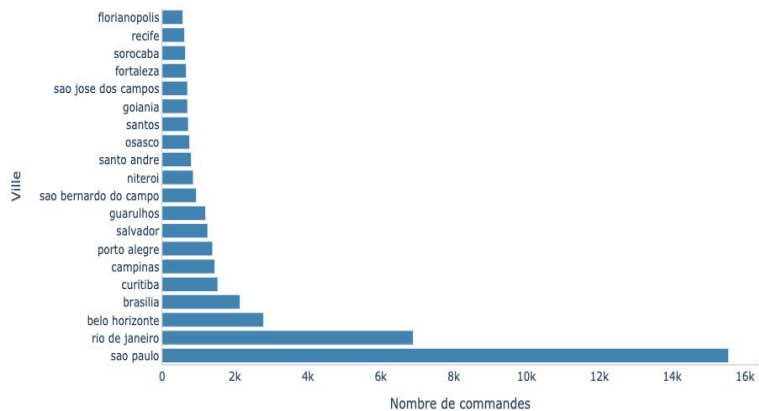
Etats Ayant Le Plus De Villes Ayant Commandé



Nombre de clients par nombre de commandes



Top 20 des villes ayant le plus de client

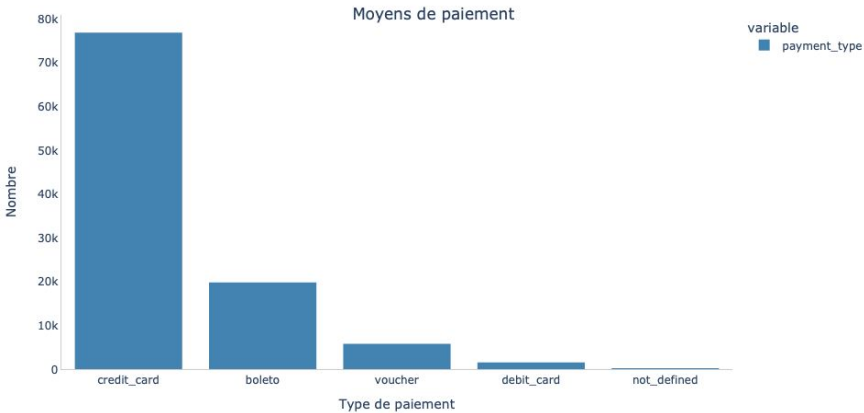


Les commandes

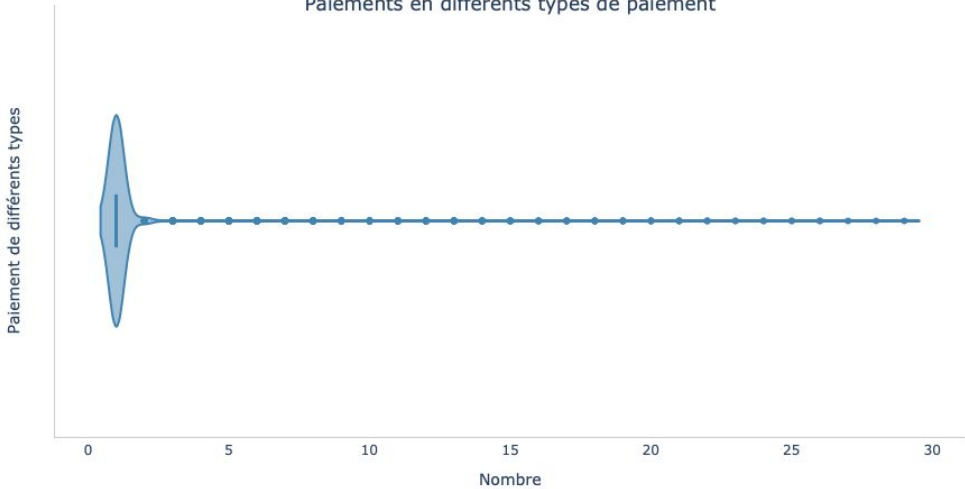
Analyse exploratoire des données



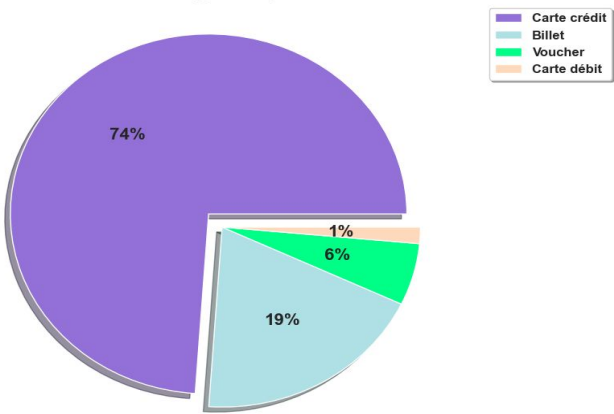
Moyens de paiement



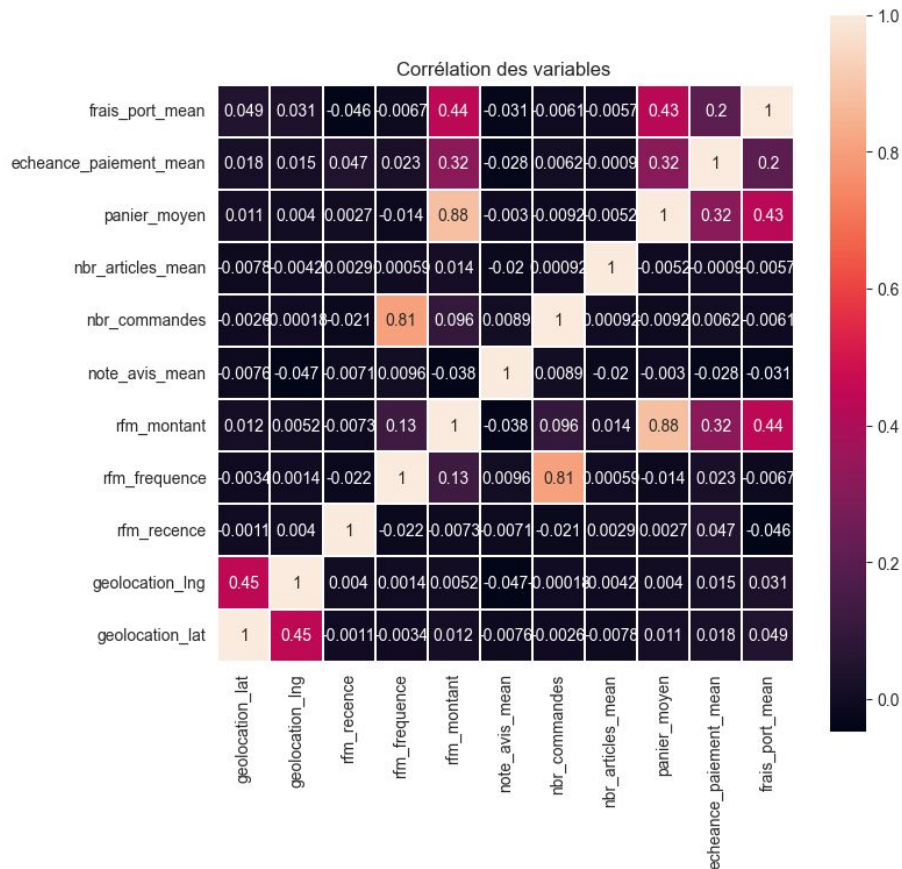
Paielements en différents types de paiement



Différents Types de paiement



Analyse exploratoire des données

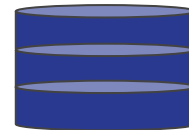


Segmentation RFM



3 Variables

Algorithmes d'apprentissages non supervisé



10 Variables

Modélisation : Segmentation

Segmentation RFM



3 Variables

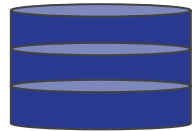


Segmentation clients

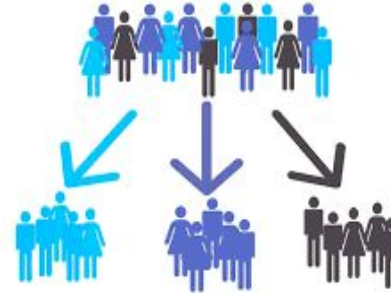


X clusters à interpréter

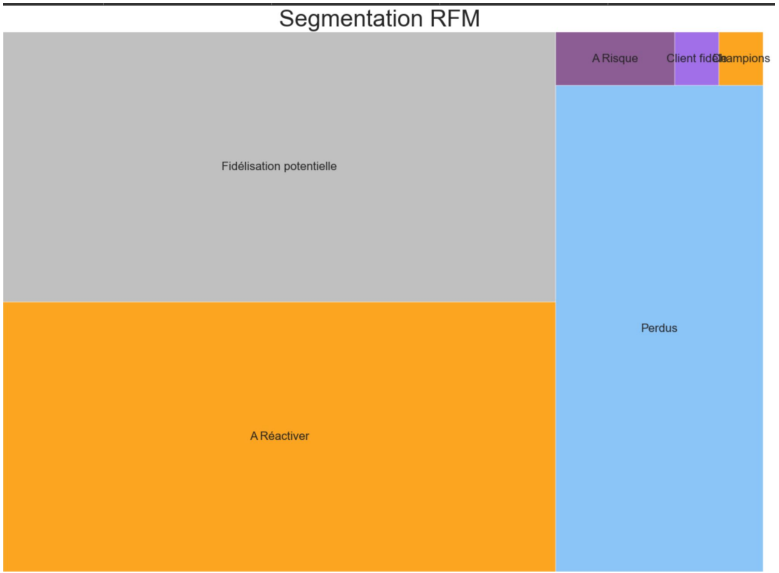
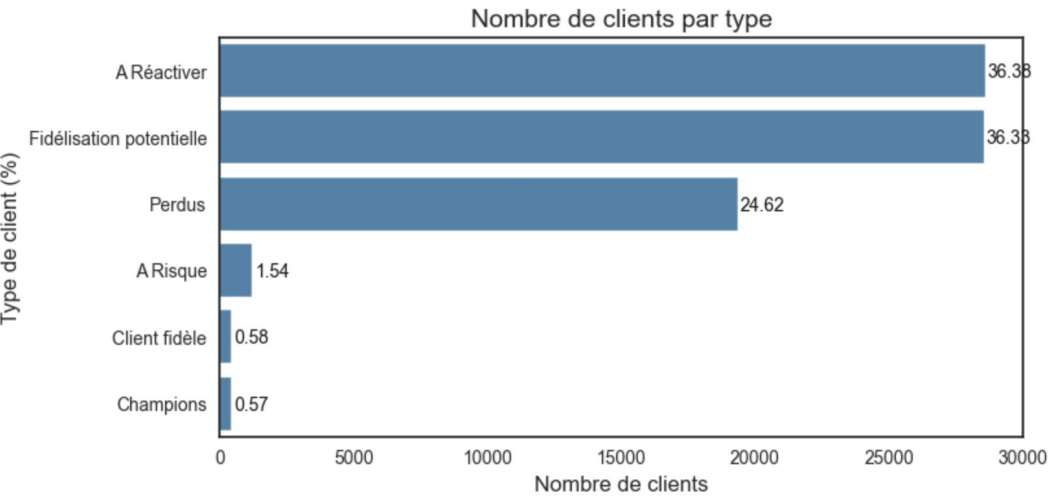
Algorithmes d'apprentissages non supervisé



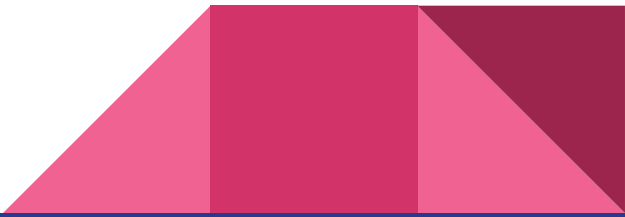
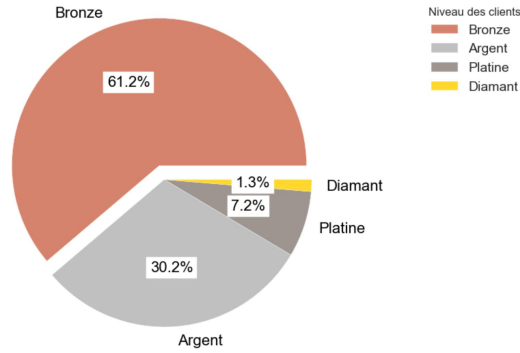
10 Variables



Modélisation: RFM Scores et Segments

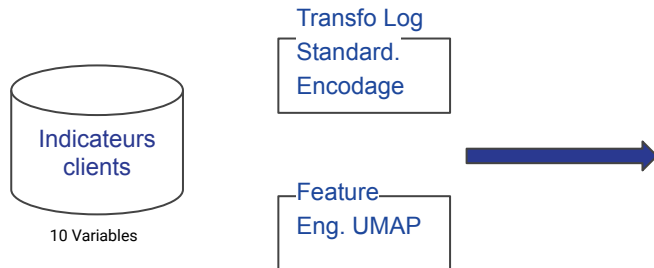


Distribution des clients par niveau diamant, platine, argent, bronze

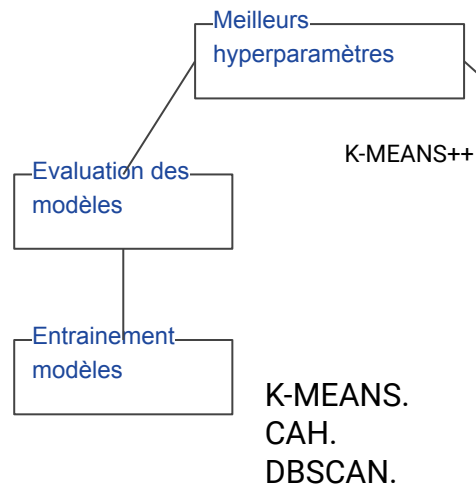


Modélisation: Clustering

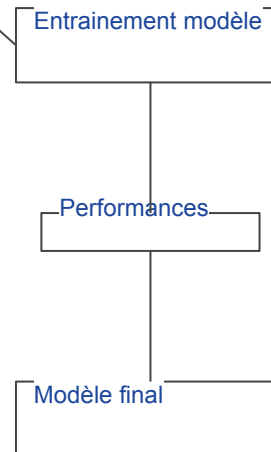
Préparation des données



Recherches d'hyper paramètres



Modélisation

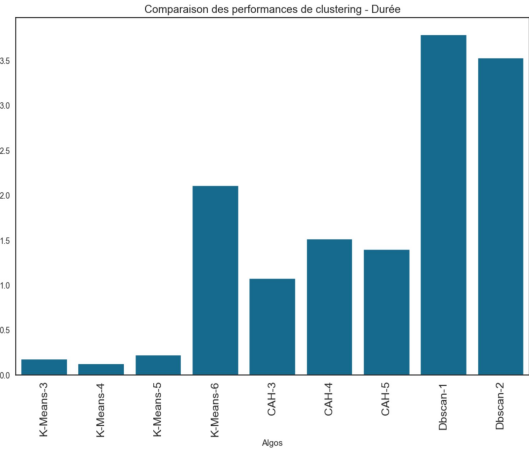
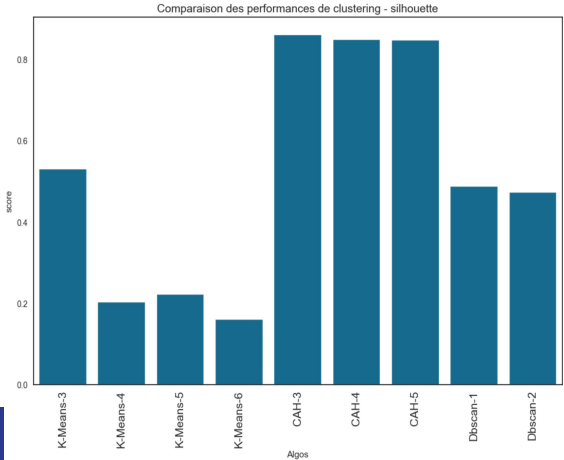
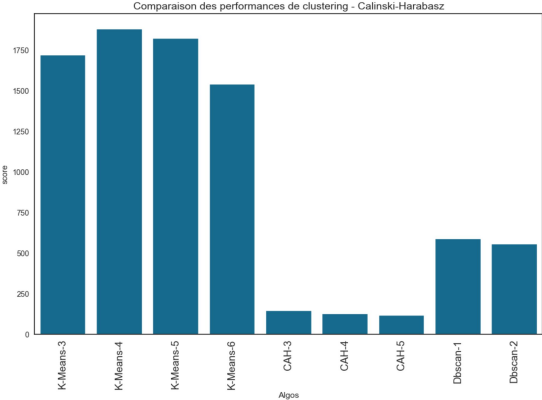
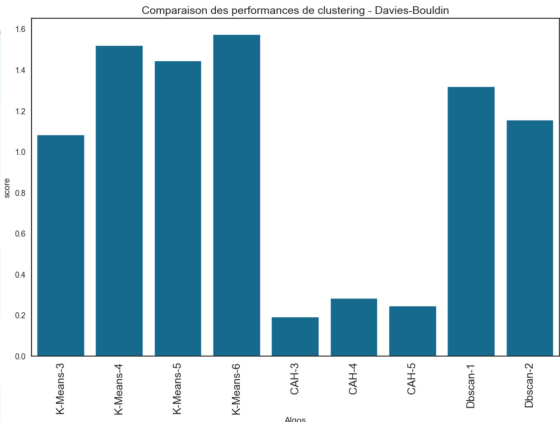


Modélisation: Performance 10000 Clients

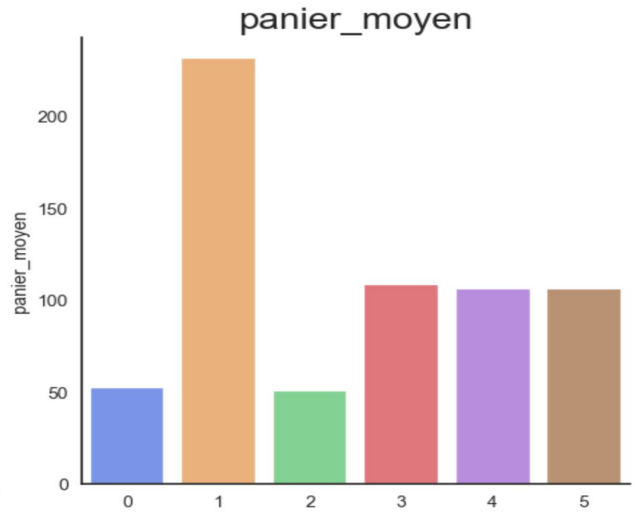
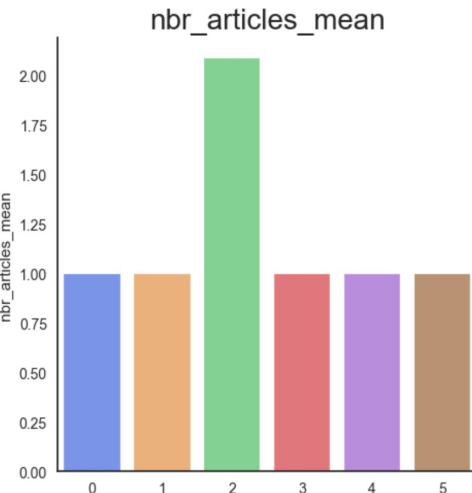
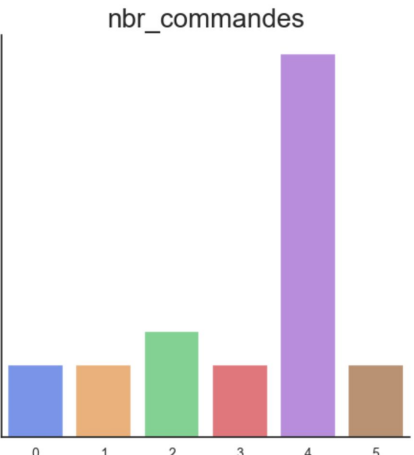
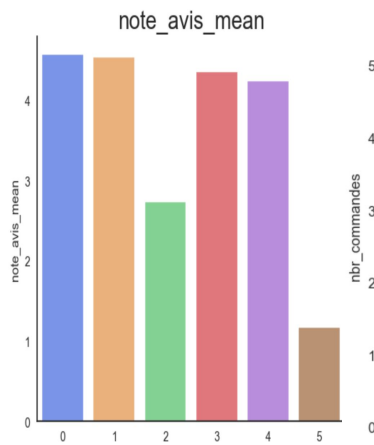
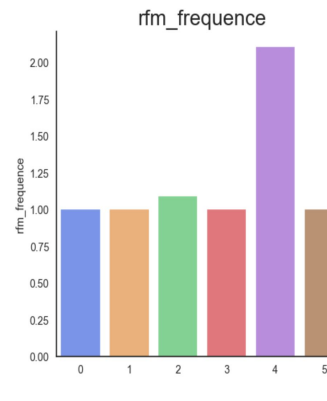
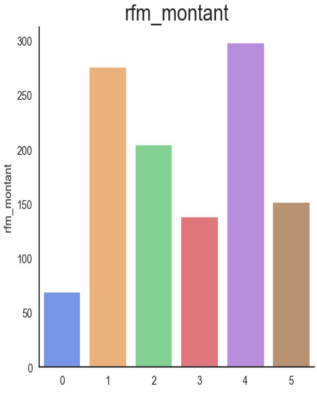
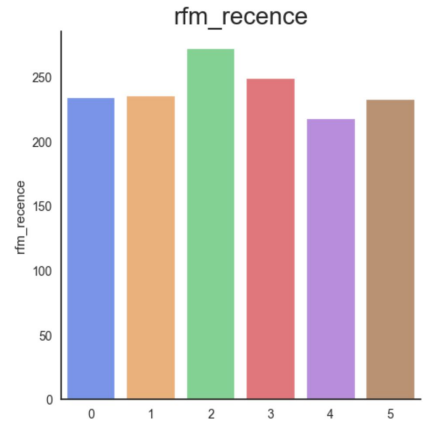
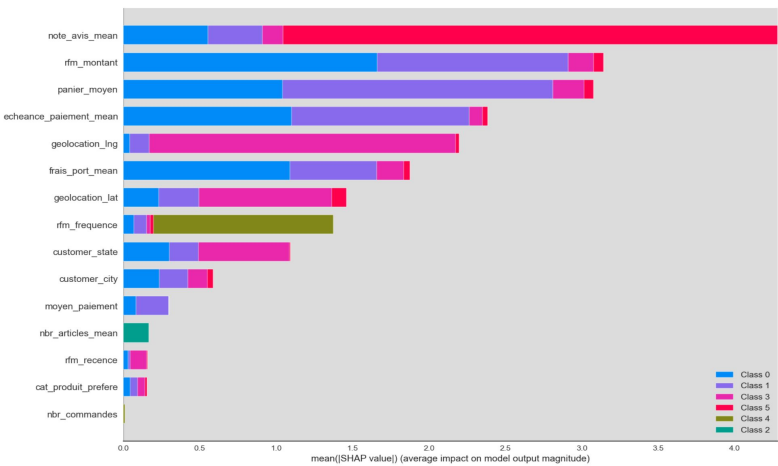


Algos	Nb_clusters	coef_silh	davies_bouldin	calinski_harabasz	Durée
K-Means-3	3	0.530003	1.082155	1719.694839	0.176034
K-Means-4	4	0.203589	1.517886	1881.768306	0.126486
K-Means-5	5	0.221475	1.443680	1821.472522	0.220118
K-Means-6	6	0.159848	1.573967	1539.842027	2.106762
CAH-3	3	0.861788	0.189780	144.356888	1.072437
CAH-4	4	0.849060	0.283107	125.487199	1.509727
CAH-5	5	0.847810	0.244944	116.614999	1.397811
Dbscan-1	4	0.487460	1.317367	587.788719	3.787756
Dbscan-2	5	0.473371	1.153617	555.013088	3.529167

Algorithme le plus performant et rapide :
K-Means

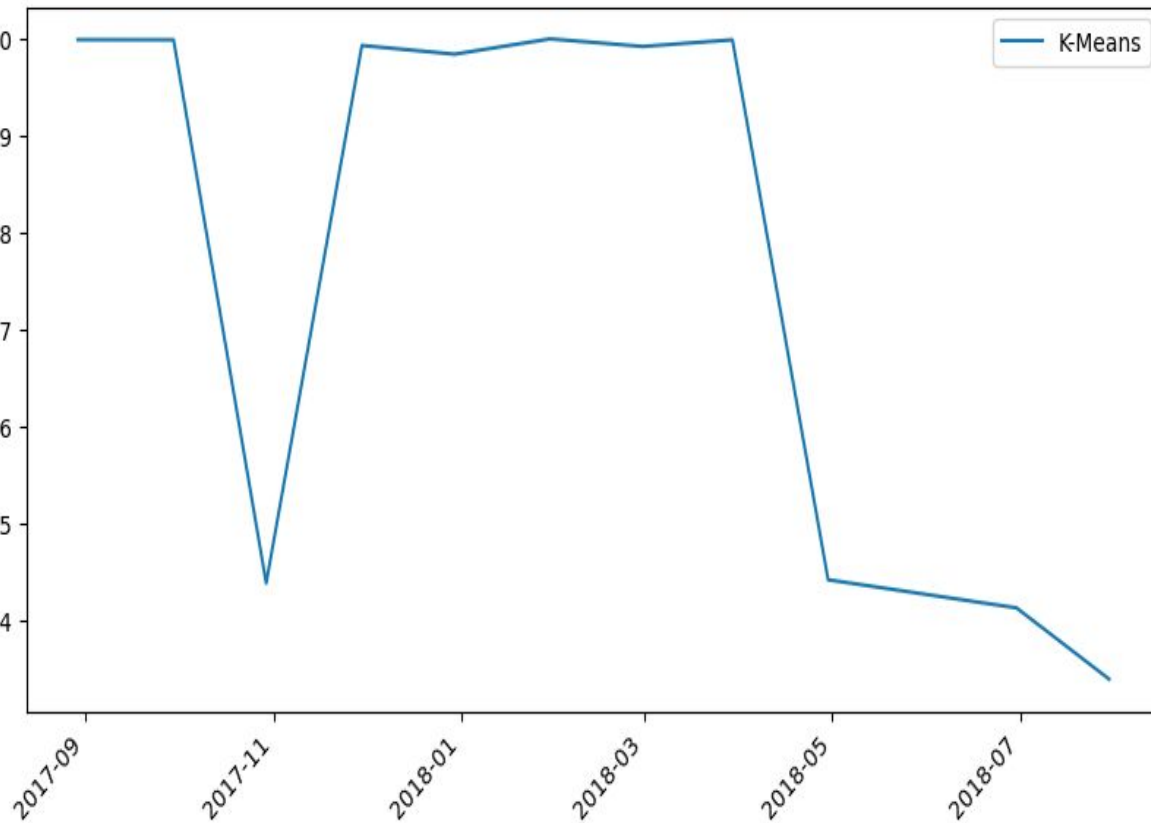


Modélisation: Clustering K-Means K= 6




Modélisation: Stabilité au cours du temps

Évolution de la stabilité des clusters KMeans sur 1 an



**Proposition d'un contrat
de maintenance avec
mise à jour trimestrielle**

PEP8

- ❖ La PEP 8 est la convention de codage officielle de Python.
 - ❖ Elle définit les règles de base pour la syntaxe, la mise en forme et le style de code en Python.
 - ❖ La PEP8 couvre des sujets tels que la mise en forme du code
 - ❖ la longueur maximale des lignes
 - ❖ l'utilisation des espaces et des tabulations
 - ❖ la nomenclature des variables et des fonctions
 - ❖ la documentation du code.
 - ❖ [En savoir plus sur PEP8](#)
- 

Conclusion et suite du projet

Modèle final :

Kmeans avec 6 clusters est une bonne piste de départ pour démarrer la segmentation de clientèle en partant d'un **contrat de maintenance avec mise à jour trimestrielle**.

Jeu de données

- Nécessite plus de données :
- démographiques (âge, profession, sexe, nombre d'enfants..)
- psychographiques (avis sur le produit, centre d'intérêt...)
- Biaisés :
- 96% des clients ne commandent qu'une seule fois.
- Notes toutes très positives.

Segmentation

- Collaborer avec l'équipe Marketing métier :
- Définir la finesse du nombre de segments souhaités par OLIST.
- Valider le choix des variables ajoutées lors du feature engineering.
- Valider les premiers résultats (modifier le paramétrage/modèle si besoin).