

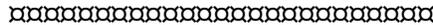
RÉPUBLIQUE DU SÉNÉGAL



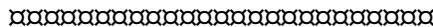
Un peuple - un but - une foi



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE
L'INNOVATION



L'excellence ma référence



Mention : Management des systèmes d'information Automatisé

Département : Économie Gestion

UFR : Science Économique Sociale



Sujet :

**La segmentation automatique des clients par machine learning
Généralité, état de l'art et étude de cas**



Présenté par :

Serigne Mor TOURE

Sous la direction de :

Dr Edouard Ngor SARR

Dr Cheikh Oumar BALDE

Sous la supervision de :

Pr Ousmane SALL

Membres du jury :

Pr Serigne DIOP (**Président**)

Dr Abel DIATTA (**Examineur 1**)

Dr Lamine FATY (**Examineur 2**)

Année universitaire : 2023 – 2024

Soutenu publiquement le 03 Aout 2024



Dédicace

Je dédie ce travail à ma mère.



Remerciements

J'adresse mes profonds remerciements :

- A mes directeurs de mémoire les Docteurs Edouard Ngor SARR Cheikh Oumar BALDE pour m'avoir confié ce travail, pour le suivi, la disponibilité, les orientations et les conseils ;
- Au Professeur Ousmane SALL pour le suivi et l'évaluation de mon mémoire. Merci à vous cher professeur ;
- Aux membres du jury qui ont accepté d'évaluer ce modeste travail ;
- Au personnel de S@RRIS GROUPE sis SALY MBOUR qui m'a bien accueilli et bien guidé pour ce travail ;
- Au corps professoral de la filière Management Informatisé des Organisations (MIO)
- A tout le corps professoral de l'Université Assane SECK de Ziguinchor pour la qualité de leur enseignement ;
- Au personnel de l'UASZ qui a rendu ce projet possible et qui continue de promouvoir l'excellence en recherche et en éducation ;
- A Serigne Moustapha FALL Fallilou;
- A Serigne Fallou DIOUCK;
- A Mama Sougou FAYE et a tous ceux de la 6c ainsi que ceux de la 5a ;
- A mes amis et ma famille pour leur soutien inébranlable, leurs encouragements et leur compréhension pendant que je me consacrais à cette étude.
- À vous tous, qui de près ou de loin, avez contribué à la réussite de ce travail.
- Vous avez toute ma gratitude.



Résumé

Ce mémoire explore l'usage des techniques d'apprentissage automatique pour segmenter les clients, facteur déterminant pour les entreprises qui aspirent à une meilleure interaction et compréhension avec leur clientèle. Ce processus implique l'analyse de grands ensembles de données afin de regrouper les clients en fonction de leurs comportements d'achat, de leurs choix préférentiels et de leurs données démographiques. Cela permet également de personnaliser les offres et les initiatives marketing, ce qui augmente la fidélité des clients et stimule les revenus. Les récents progrès dans les nouvelles techniques d'apprentissage automatique ont considérablement amélioré la précision des modèles de segmentation client, ouvrant ainsi de nouveaux horizons dans des domaines variés comme le commerce électronique, la finance, la santé, l'éducation et les télécommunications. À travers une étude de cas pratique, ce mémoire démontre comment une entreprise peut mettre en œuvre efficacement le machine learning pour segmenter sa clientèle, adapter ses stratégies marketing et améliorer ses performances commerciales.

Mots-clés :

Apprentissage automatique, Analyse de données, Segmentation client, Marketing, Données.



Abstract

This thesis explores the use of machine learning for customer segmentation, a crucial factor for companies aiming to better interact with and understand their clientele. This process involves analyzing large datasets to group customers based on their purchasing behaviors, preferences, and demographic data. It also enables the personalization of offers and marketing initiatives, which increases customer loyalty and boosts revenue. Recent advancements in new machine learning techniques have significantly improved the accuracy of customer segmentation models, opening new opportunities in various fields such as e-commerce, finance, healthcare, education, and telecommunications. Through a practical case study, this thesis demonstrates how a company can effectively implement machine learning to segment its customer base, tailor its marketing strategies, and enhance its business performance.

Keywords:

Machine Learning, Data Analysis, Customer Segmentation, Marketing, E-commerce & Data.



Sommaire

Dédicace	i
Remerciements	ii
Résumé	iii
Sommaire.....	v
Liste des figures.....	vi
Liste des tableaux	viii
Sigles et abréviations.....	ix
Introduction générale.....	1
Chapitre 1 : Généralité sur la segmentation des clients et le machine Learning	3
Chapitre 2 : Le Machine Learning et la Segmentation des clients	22
Chapitre 3 : Etat de l'art et positionnement.....	30
Chapitre 4 : Etude de cas	36
Conclusion et Perspectives	58
Références	59
Table des matières	70
Annexe 1 : Segmentation données clients	73
Annexe 2 : prédiction données clients.....	75
Annexe 3 : Segmentation données transactionnelles	79
Annexe 4 : prédiction données transactionnelles	81



Liste des figures

Figure 1: Exemple de segments RFM utilisés par Octolis [w29].....	8
Figure 2 : Les types de machines Learning [7]	10
Figure 3: L'apprentissage supervisé [7].....	11
Figure 4: L'apprentissage non supervisé [7].....	12
Figure 5: L'apprentissage par renforcement [7]	13
Figure 6: Différence entre le Machine Learning et le Deep Learning [w28].....	14
Figure 7 option de segmentation dans ibm Watson studio [w34]	33
Figure 8: Graphique de la segmentation des clients par âge.	45
Figure 9: Graphique de la segmentation des clients par Genre.	45
Figure 10: Graphique de la segmentation des clients par Etat.	46
Figure 11: Graphique de la segmentation des clients par âge et genre.	47
Figure 12: Graphique de la segmentation des clients par âge et Etat.	47
Figure 13: Graphique de la segmentation des clients par genre et Etat.	48
Figure 14: Graphique de la segmentation des clients par âge, genre et Etat.	49
Figure 15: Graphique de la prédiction du Genre du prochain client.	50
Figure 16: Graphique de la prédiction de l'âge du prochain client.	50
Figure 17: Graphique de la prédiction de l'Etat du prochain client.	51
Figure 18: Graphique de la prédiction de l'âge, Etat et le genre du prochain client.	52
Figure 19: Graphique de la segmentation des transactions par Marque.....	53
Figure 20: Graphique de la segmentation des transactions par commande en ligne.....	54
Figure 21: Graphique de la segmentation des transactions par Prix de vente et commande en ligne.....	54
Figure 22: Graphique de la segmentation des transactions par marque et commande en ligne.	55
Figure 23: Graphique de la segmentation des transactions par marque et prix de vente.	55
Figure 24: segmentation des transactions par commande en ligne et prix de vente	56
Figure 25: Graphique de la segmentation des transactions par marque, commande en ligne et prix de vente.	57
Figure 26: Graphique de la prédiction de la prochaine marque qui sera achetée par un client.	54
Figure 27: Graphique de la prédiction de la prochaine commande en ligne du client.	55
Figure 28: Prédiction du prochain prix de vente du prochain client	56



Figure 29: Graphique de la prédiction de la marque, de la commande en ligne et du prix de vente de la prochaine transaction. 57



Liste des tableaux

Tableau 1: Tableau contenant les algorithmes de ML les plus utilisées	15
Tableau 2:Présentation des données Dataset.....	42
Tableau 3: Outils choisis	43
Tableau 4: Guide d'installation des outils et API.....	43



Sigles et abréviations

ACP : Analyse en Composantes Principales (Principal Component Analysis en anglais)

AFD : Analyse Factorielle Discriminante

API : Application Programming Interface (Interface de Programmation Applicative)

ASR : Automatic Speech Recognition (Reconnaissance Automatique de la Parole)

BIRCH: Balanced Iterative Reducing and Clustering using Hierarchies

CDP : Customer Data Platform (Plateforme de Données Client)

CNN : Convolutional Neural Network (Réseau de Neurones Convolutif)

CSV : Comma-Separated Values (Valeurs Séparées par des Virgules)

DL: Deep Learning

DQN: Deep Q-Network

GMM : Gaussian Mixture Model (Mélange de Modèles Gaussiens)

HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise

IA : Intelligence Artificielle (Artificial Intelligence en anglais)

IBM: International Business Machines Corporation

IoT : Internet of Things (Internet des Objets)

JSON : JavaScript Object Notation (Notation Objet JavaScript)

k-NN: k-Nearest Neighbors (k Plus Proches Voisins)

Mask R-CNN: Mask Region Convolutional Neural Network

NMF : Non-negative Matrix Factorization (Factorisation de Matrices Non Négatives)

NLP : Natural Language Processing (Traitement du Langage Naturel)

NLU : Natural Language Understanding (Compréhension du Langage Naturel)

OPTICS : Ordre Préférentiel et Indexation de Clustering Spatial

RFM: Recency, Frequency, Monetary (Réponse, Fréquence, Montant)

RNN : Recurrent Neural Network (Réseau de Neurones Récurrent)

SARSA: State-Action-Reward-State-Action

SOM: Self-Organizing Map (Carte Auto-Organisatrice)

SQL : Structured Query Language (Langage de Requête Structurée)

VADER: Valence Aware Dictionary and sEntiment Reasoner

VIP : Valeur d'Importance Perçue

XGBoost: Extreme Gradient Boosting

XML : eXtensible Markup Language (Langage de Balisage Extensible)

YOLO: You Only Look Once



Introduction Générale

Dans un monde où les entreprises veulent constamment mieux comprendre ce que veulent les clients, la segmentation de ces derniers devient une stratégie très importante. En effet, la segmentation des clients consiste à diviser la base de clients en groupes homogènes selon des critères communs tels que les caractéristiques démographiques, les habitudes d'achat, les préférences ou les comportements de navigation, afin de mieux cibler les stratégies marketing et répondre aux attentes spécifiques de chaque segment. Son but est de comprendre et de répondre aux différentes attentes et préférences des diverses cohortes de clients. En personnalisant les produits, les services et les efforts de marketing en fonction de ces segments, les entreprises peuvent accroître la satisfaction de la clientèle, augmenter les ventes et optimiser la performance globale des stratégies marketing [W7].

Mais face à la complexité croissante des comportements d'achat et à la diversité des clients, les entreprises cherchent des solutions pour mieux comprendre leur marché. Dans cette optique, elles font alors recours à de nouvelles méthodes et outils de segmentation. Un des outils les plus en vogue aujourd'hui est le machine learning. Ainsi, *comment l'usage du machine learning peut-il aider les entreprises à améliorer la segmentation des clients et à approfondir leur compréhension du marché ?* En d'autres termes :

- Les algorithmes de machine learning peuvent-ils aider à ressortir des segments plus précis pour optimiser les choix stratégiques ?
- Pouvons-nous aspirer à de meilleures prédictions des opérations et des groupes de clients afin d'anticiper les actions à faire ?

Nous pensons ainsi que, grâce au machine learning, la segmentation va considérablement s'améliorer et offrir de nouvelles possibilités pour segmenter les clients de manière plus précise et adapter à chacun tout en offrant des prédictions claires.

L'objectif principal de cette étude est d'explorer les multiples aspects de la segmentation des clients par le machine learning, en examinant à la fois les bases théoriques, les avancées technologiques et les applications pratiques dans différents domaines. En d'autres termes, l'objectif est d'utiliser des techniques de machine learning pour examiner le fonctionnement de la segmentation des clients et évaluer son efficacité dans le contexte actuel des entreprises. Pour atteindre cet objectif général, nous nous focaliserons sur trois objectifs spécifiques à savoir :

- Analyser les concepts et méthodes de base de la segmentation des clients par machine learning ;
- Étudier les avancées récentes en passant par la revue la littérature existante ;



- Illustrer l'application pratique du machine learning dans la segmentation des clients avec une étude de cas concrète.

Cette recherche adoptera une approche mixte, combinant l'approche quantitative et celle qualitative à travers une revue de la littérature détaillée et une étude de cas approfondie dans un secteur spécifique, tel qu'une entreprise commerciale.

Notre travail sera organisé en quatre chapitres :

- **Chapitre 1 : Généralités sur la segmentation des clients** : Nous explorerons les bases de la segmentation des clients, en définissant son importance pour les entreprises et en examinant ses applications dans différents secteurs. Aussi, nous nous concentrerons sur les principes fondamentaux du machine learning, en expliquant son fonctionnement ;
- **Chapitre 2 : Le machine learning et segmentation** : Nous explorons l'utilité et les cas d'utilisation du machine learning dans la segmentation des clients. Des exemples concrets viendront éclairer ces explications ;
- **Chapitre 3 : État de l'art** : Nous passerons en revue les dernières recherches et pratiques dans le domaine de la segmentation des clients et du machine learning. Cette revue inclura les tendances, méthodes et outils les plus récents utilisés dans ce domaine en évolution ;
- **Chapitre 4 : Étude de cas** : Nous présenterons une étude de cas concrète où nous appliquerons les concepts discutés dans les chapitres précédents. Nous détaillerons la méthodologie utilisée, présenterons nos résultats et partagerons les leçons apprises au cours de cette expérience.



Chapitre 1 : Généralité sur la segmentation des clients et le machine Learning



1.1. Données clients

1.1.1. Définitions

Les données clients sont des informations recueillies sur les clients d'une entreprise, telles que leur nom, adresse, âge, préférences d'achat, historique d'achats et comportements en ligne. Elles sont obtenues lors des interactions entre le client et l'entreprise par le biais de divers moyens de communication tels qu'un site internet, des applications mobiles, des sondages téléphoniques, les réseaux sociaux, des campagnes de marketing ou d'autres canaux en ligne et hors ligne. Ces données [1] sont essentielles pour toute stratégie d'entreprise, car elles fournissent des informations sur les comportements, les caractéristiques démographiques et les préférences des clients. Elles aident les entreprises à mieux comprendre leurs clients, à communiquer efficacement avec eux et à les fidéliser. En outre, elles offrent un aperçu précieux des attentes des clients vis-à-vis des marques, des produits qu'ils recherchent et de leurs préférences en termes d'interaction [W1].

1.1.2. Types de données

Entre la plateforme et les clients plusieurs échanges et création de données existent. Ainsi nous avons une diversité de données dont certaines n'entrent pas dans le cadre de notre étude.

Pour notre étude, nous nous intéressons aux :

- **Données personnelles ou de profils** : Ce sont toutes les informations permettant d'identifier une personne physique, comme le nom, le numéro de téléphone, l'adresse e-mail, l'âge, le sexe, le numéro de sécurité sociale, les données biométriques et l'adresse résidentielle. Elles sont collectées pour des raisons de communication, d'interaction ou d'identification [W2]. L'utilisation de ces données est cruciale dans divers secteurs d'activité comme le secteur médical [2], financier etc. ;
- **Données géographiques** : également appelées données spatiales ou données géospatiales identifient l'emplacement géographique des entités. Ces données incluent tout ce qui peut être associé à un emplacement sur le globe, ou plus simplement tout ce qui peut être cartographié. Par exemple, les routes, les frontières des pays, localisation(mobile), GPS, des adresses (physiques), des cartes (topographiques), et d'autres données liées à la position sur la Terre, sont tous des types de données spatiales [w3]. Aujourd'hui, elles sont souvent désignées par "spatiales" ou "géospatiales" pour éviter la confusion entre la représentation cartographique et la nature géographique [3] ;
- **Données transactionnelles** : Les données transactionnelles sont des informations enregistrées lors des transactions, comprenant l'heure et le lieu de l'achat, les prix des articles, le mode de paiement, les éventuelles réductions, ainsi que d'autres détails



pertinents. Elles sont généralement collectées au moment de l'achat [w4] et décrivant les activités et échanges entre les concernés.

1.1.3. Les sources de données clients

Une source de données désigne le lieu physique ou virtuel où les informations sont stockées dans divers formats ; en somme, c'est là d'où viennent les données. Elle peut être un ordinateur, une base de données, un fichier, un capteur, un appareil IoT, une API, ou toute autre entité pouvant transmettre des données à des applications ou des processus [w5]. Parmi celle-ci on peut citer :

- Les fichiers journaux (logs) ;
- Les fichiers de données structurées (CSV, JSON, XML, etc.) ;
- Les fichiers de bases de données (SQL, SQLite, etc.) ;
- Les fichiers de feuilles de calcul (Excel, Google Sheets, etc.) ;
- Les fichiers de texte brut ;
- Les fichiers de configuration (INI, YAML, etc.) ;
- Les fichiers de rapports générés automatiquement ;
- Les fichiers de sauvegarde et de récupération ;
- Les fichiers de documentation technique ;
- Les fichiers de résultats de tests et d'analyses ;
- Etc.

Ces sources de données sont couramment utilisées dans divers environnements pour stocker, échanger et analyser des informations structurées et non structurées.

Dans le cas de notre étude, nous intéresserons exclusivement sur les formats de données Excel et CSV.

1.2. La segmentation des clients

1.2.1. Définition

La segmentation du client consiste à diviser la base de clients en groupes homogènes selon des critères communs tels que les caractéristiques démographiques, les habitudes d'achat, les préférences ou les comportements de navigation, afin de mieux cibler les stratégies marketing et répondre aux attentes spécifiques de chaque segment. Son but est de comprendre et de répondre aux différentes attentes et préférences des diverses cohortes de clients. En personnalisant les produits, les services et les efforts de marketing en fonction de ces segments, les entreprises peuvent accroître la satisfaction de la clientèle, augmenter les ventes et optimiser la performance globale des stratégies marketing [W7].



1.2.2. Les rôles de la segmentation des clients

Un segment est considéré comme opérationnel lorsqu'il peut être ciblé de manière spécifique. La taille des segments sélectionnés doit être suffisante pour assurer la rentabilité d'une approche dédiée. En effet, la segmentation permet de mettre en œuvre des actions marketing différenciées selon les segments et de proposer, si nécessaire, une offre de produits spécifique à chacun, tout en évitant les conflits entre segments. Il est également possible de choisir de ne cibler qu'un seul segment de marché [6]. Elle est aussi un outil stratégique essentiel qui permet de répondre de manière plus précise aux besoins des clients, d'améliorer l'efficacité du marketing, de renforcer la fidélité et d'identifier de nouvelles opportunités de croissance, tout en optimisant l'utilisation des ressources et en minimisant les risques. Ainsi, le rôle de la segmentation des clients est crucial pour :

- Adapter les offres et les communications afin de répondre de manière précise aux besoins particuliers de chaque segment de clientèle ;
- Diriger de manière ciblée les campagnes marketing en envoyant des messages spécifiquement adaptés à chaque segment de la clientèle ;
- Optimiser l'utilisation des ressources marketing en focalisant les efforts sur les segments de clientèle les plus rentables ou présentant un fort potentiel de croissance ;
- Accroître la satisfaction des clients en proposant des produits et des services adaptés à leurs besoins individuels ;
- Détecter des possibilités de croissance sur le marché en identifiant de nouveaux segments ou des besoins non comblés au sein des segments existants ;
- Diminuer les risques en élargissant la base de clients afin d'éviter de dépendre excessivement d'un seul segment ;
- Améliorer l'efficacité des campagnes marketing en classant les clients en fonction de leurs comportements d'achat ou de leurs préférences ;
- Renforcer la fidélisation des clients en offrant des programmes de fidélité spécifiquement adaptés à chaque segment de clientèle ;
- Créer des produits ou des services sur mesure, en tenant compte des besoins et des spécificités de chaque segment de clientèle ;
- Renforcer la compétitivité en évaluant les atouts et les faiblesses de chaque segment par rapport à la concurrence ;
- Evaluer l'efficacité des diverses stratégies marketing en examinant les réactions des différents segments de clientèle ;
- Réajuster les stratégies de tarification selon la réactivité aux prix de chaque groupe ;



- Améliorer l'expérience client en adaptant les interactions selon les préférences de chaque groupe ;
- Diminuer le gaspillage en évitant d'allouer des ressources aux segments moins appropriés.
- Anticiper les actions à venir des clients en examinant les tendances propres à chaque groupe ;
- Segmenter les clients en fonction de leurs besoins en matière de support ou de service après-vente ;
- Personnaliser les recommandations de produits ou services en fonction des habitudes d'achat de chaque segment ;
- Optimiser la localisation des points de vente ou des canaux de distribution en fonction des préférences géographiques de chaque segment ;
- Adapter les politiques de retour et de garantie en fonction des attentes spécifiques de chaque segment ;
- Maximiser la rentabilité en identifiant les segments de clients les plus lucratifs ;
- Faciliter l'innovation en découvrant de nouvelles opportunités basées sur les besoins spécifiques des segments ;
- Planifier plus efficacement les inventaires et les stocks en fonction des prévisions de la demande de chaque segment ;
- Renforcer la communication interservices en fournissant des données claires sur les attentes de chaque segment.

1.2.3. Les méthodes de la segmentation des clients

Il existe différentes méthodes pour diviser la clientèle en segments distincts dont :

- **La segmentation Démographique** : c'est le fait de diviser les clients en groupes basés sur des caractéristiques démographiques. En effet, ces caractéristiques peuvent être l'âge, le genre, le revenu, l'état civil, le niveau d'éducation et la profession. Par exemple, segmenter les clients en fonction de leur âge permet de cibler les jeunes adultes avec des publicités de produits technologiques et les seniors avec des offres de services de santé ;
- **La segmentation géographique** : c'est le fait de classer les clients selon leur localisation, comme le pays, la région etc. En effet, ces critères peuvent inclure divers aspects géographiques tels que la ville, le climat et la distinction entre zones urbaines et rurales. Par exemple, une entreprise de vêtements peut proposer des vêtements d'hiver aux clients des régions froides et des vêtements légers aux clients des régions chaudes ;



- **La segmentation psychographique** implique la classification des clients en groupes en fonction de leurs modes de vie, opinions, personnalités, valeurs et intérêts. Par exemple, les amateurs de sports extrêmes peuvent être ciblés avec des équipements spécialisés et des voyages d'aventure ;
- **La segmentation comportementale** classe les clients en fonction de leurs comportements d'achat, de leurs habitudes de consommation et de leur fidélité à la marque, incluant la fréquence et les occasions d'achat, la réponse aux promotions et l'utilisation du produit. Par exemple, les clients réguliers peuvent bénéficier de remises spéciales et de programmes de fidélité pour les inciter à acheter plus fréquemment ;

Description des segments			
Segment	R	F*M	Description
Champions	5,6	5,6	Ont commandé récemment, dépensent souvent et le plus.
Loyaux	3,4,5,6	4,5,6	Ont plusieurs commandes dernièrement.
Loyalistes potentiels	4,5,6	2,3,4	Clients relativement nouveau, mais avec de bon paniers et plus d'une commande.
Récents	5,6	1,2	Ont commandé récemment, mais peu souvent.
Prometteurs	4,5	1,2	Ont commandé récemment, mais des montants moyens.
A surveiller	3,4	3,4	Récence, fréquence et montant légèrement au-dessus de la moyenne.
A réactiver	3,4	1,2,3	Récence, fréquence et montant légèrement en-dessous de la moyenne, à réactiver.
A risque	1,2,3	3,4,5,6	Ont commandé beaucoup et souvent mais il y a quelque temps, il faut les faire revenir !
A ne pas perdre	1,2	5,6	Ont commandé beaucoup et souvent mais ne sont pas revenus depuis longtemps.
En sommeil	2,3	2,3	N'ont pas commandé depuis longtemps, montant et fréquence faibles.
Perdus	1,2,3	1,2,3	N'ont pas commandé depuis longtemps, montant et fréquence les plus faibles.

Figure 1: Exemple de segments RFM utilisés par Octolis [w29].

- **La segmentation par Bénéfices Recherchés** divise les clients en groupes en fonction des avantages spécifiques qu'ils recherchent dans un produit ou un service, tels que la qualité du produit, le prix, le service client, la commodité et les fonctionnalités spécifiques. Par exemple, les clients qui accordent de l'importance à la qualité peuvent être attirés par des campagnes mettant en avant la durabilité et la performance du produit ;
- **La segmentation par Valeur Client** identifie les clients en fonction de leur valeur à vie et de leur rentabilité pour l'entreprise, incluant le coût d'acquisition du client. Par exemple, les clients à forte valeur peuvent recevoir un service client premium et des offres exclusives pour les fidéliser ;
- **La segmentation par Canaux de Communication** distingue les clients en fonction de leurs préférences pour les différents moyens de communication tels que les e-mails, les SMS, les réseaux sociaux ou les appels téléphoniques. Par exemple, les clients préférant



les e-mails peuvent recevoir des newsletters et des offres par e-mail, tandis que ceux préférant les SMS peuvent recevoir des alertes de promotions par SMS ;

- **La Segmentation Technologique** classe les clients selon leur utilisation des appareils (mobile, tablette, desktop) et leur compétence technologique, incluant l'adoption des nouvelles technologies. Par exemple, les utilisateurs de smartphones peuvent être ciblés avec des applications mobiles et des offres de produits compatibles avec les technologies mobiles ;
- **La segmentation par Étape du Cycle de Vie du Client** regroupe les clients en fonction de leur stade dans le cycle d'achat, tels que les nouveaux clients, les clients réguliers, les clients fidèles et les clients inactifs. Par exemple, les nouveaux clients peuvent recevoir des offres de bienvenue, tandis que les clients fidèles peuvent bénéficier de programmes de fidélité ;
- **La segmentation par Secteur d'Activité (B2B)¹** classe les entreprises clientes en fonction de leur industrie, des types de produits ou services achetés, de leur taille et de leur position dans la chaîne de valeur. Par exemple, une entreprise de logiciels peut offrir des solutions spécifiques pour les secteurs de la santé, de la finance et de l'éducation ;
- **La segmentation par Besoins et Attentes** divise les clients en groupes en fonction de leurs besoins fonctionnels et émotionnels, ainsi que de leurs attentes en matière de service. Par exemple les clients recherchant un service rapide peuvent être ciblés avec des options de livraison express et des services après-vente réactifs ;
- **La segmentation par Situation d'Achat** classe les clients en fonction du contexte dans lequel ils effectuent leurs achats, comme les achats de routine, impulsifs ou pour des occasions spéciales, tels que l'achat de solution à un problème ou pour des événements particuliers. Par exemple, les clients achetant pour des occasions spéciales comme des anniversaires peuvent être ciblés avec des offres de cadeaux et des emballages personnalisés.

1.3. Le machine learning

1.3.1. Définitions et fonctionnement

Le machine learning, ou apprentissage automatique en français, est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur le développement de systèmes informatiques capables d'apprendre et de s'améliorer sans être explicitement programmés. En effet, c'est une

¹ B2B, abréviation de "Business to Business", se réfère aux transactions commerciales entre entreprises.

étude scientifique des algorithmes et des modèles statistiques utilisés par les systèmes informatiques pour effectuer une tâche spécifique sans être explicitement programmés [7]. Plutôt que de suivre des instructions programmées, les modèles de machine learning utilisent des algorithmes pour analyser des données et tirer des conclusions, ce qui leur permet de prendre des décisions ou de réaliser des tâches sans être spécifiquement programmés pour chaque étape [w8]. En d'autres termes, le Machine Learning est le processus qui consiste à entraîner un programme, appelé modèle, pour effectuer des prédictions utiles à partir de données [w35].

1.3.2. Les types de machine learning

Le machine learning est un champ assez vaste, et il est utilisé partout aujourd'hui que ce soit dans la santé, la finance, le commerce électronique, l'industrie, le transport, les sciences sociales, l'éducation, les ressources humaines et surtout sur le domaine marketing. On distingue principalement quatre techniques de Machine Learning qui sont : l'apprentissage supervisé, l'apprentissage non-supervisé, l'apprentissage semi-supervisé, et l'apprentissage par renforcement. Dans l'image ci-dessous on a catégorisé l'apprentissage automatique en quatre types : supervisé, semi-supervisé, non supervisé et par renforcement, chacun ayant des méthodes et des objectifs distincts.

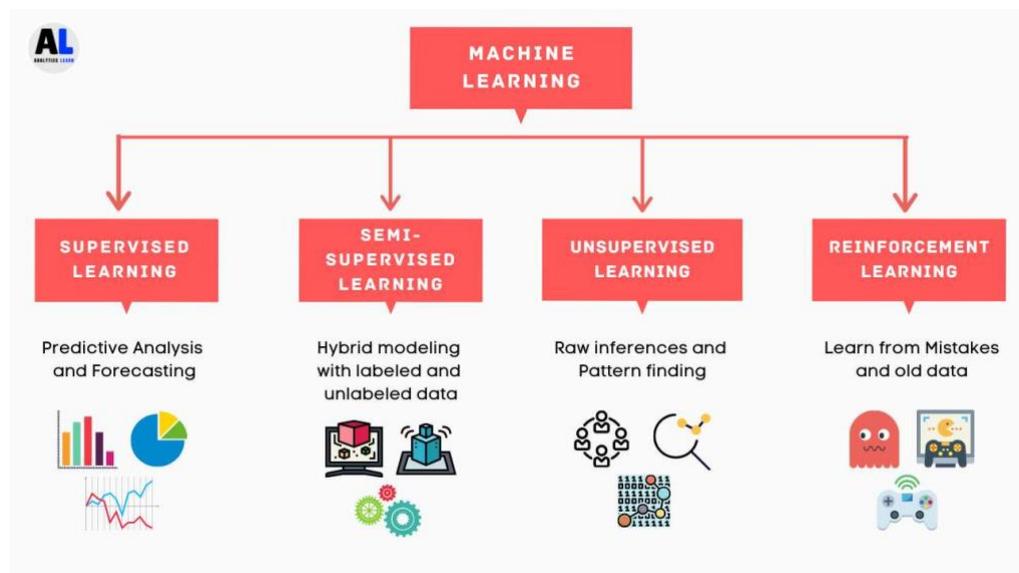


Figure 2 : Les types de machines Learning [w45]

1.3.2.1. L'apprentissage supervisé

L'apprentissage supervisé est le paradigme d'apprentissage le plus populaire en Machine Learning. Comme son nom l'indique, cela consiste à superviser l'apprentissage de la machine en lui montrant des exemples de la tâche qu'elle doit réaliser. En effet, elle utilise des données

étiquetées pour entraîner des algorithmes à classer des données ou à prédire des résultats, en ajustant ses pondérations grâce à la validation croisée pour résoudre divers problèmes réels à grande échelle[w9]. Il existe plusieurs algorithmes d'apprentissage automatique supervisé, comme les machines à vecteur de support, les algorithmes de régression linéaire et logistique et les forêts aléatoires [w10]. Les tâches courantes de l'apprentissage supervisé comprennent la classification (attribuer une catégorie ou une classe à une entrée, ou chercher à le prédire) et la régression (prédire une valeur continue ou une quantité). Les algorithmes populaires incluent la régression linéaire, les arbres de décision, les réseaux de neurones et les machines à vecteurs de support (SVM) [w11]. Ci-dessous on a une image qui représente le processus de développement et de déploiement d'un modèle d'apprentissage automatique, en commençant par la collecte de données, l'entraînement du modèle, l'évaluation, l'ajustement, et enfin le déploiement en production.

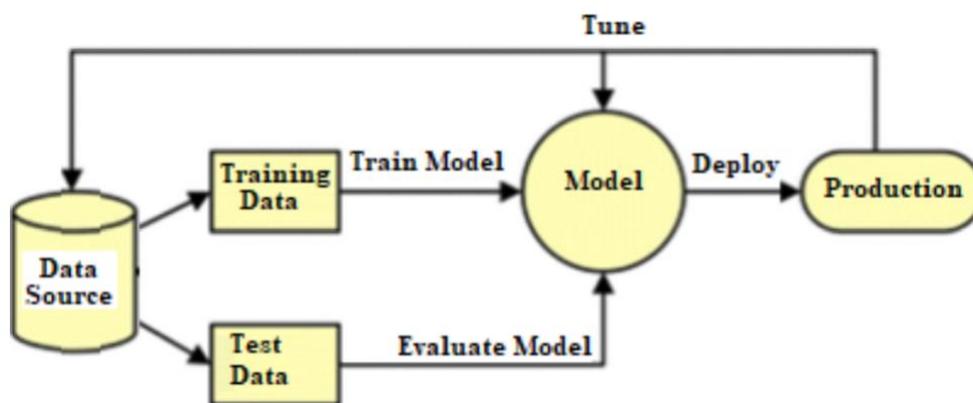


Figure 3: L'apprentissage supervise [7]

L'image présente un schéma de flux pour le développement de modèles d'apprentissage automatique, incluant la séparation des données en ensembles d'entraînement et de test, l'entraînement du modèle, son évaluation et ajustement, puis son déploiement en production après validation.

1.3.2.2.L'apprentissage non supervisé

Les algorithmes d'apprentissage non supervisé explorent les données de manière autonome pour en découvrir la structure. En effet, ils identifient des caractéristiques clés des données et, lorsqu'ils reçoivent de nouvelles informations, utilisent ces caractéristiques pour classer les données. Cet apprentissage est principalement utilisé pour le regroupement et la réduction des dimensions des caractéristiques [7]. Les algorithmes utilisés sont essentiellement dédiés au



clustering [w10]. Les tâches courantes de l'apprentissage non supervisé comprennent le regroupement (grouper des entrées similaires) et la réduction de la dimensionnalité (simplifier la représentation des données). Les algorithmes populaires incluent K-means, l'analyse en composantes principales (ACP), les méthodes de regroupement hiérarchique et la réduction de la dimensionnalité et les autoencodeurs [w11]. Ci-dessous, on a une image qui illustre le concept de l'apprentissage non supervisé, où des algorithmes apprennent à partir de données brutes sans étiquettes pour découvrir des structures cachées.

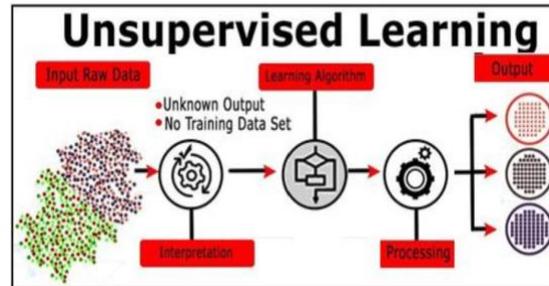


Figure 4: L'apprentissage non supervise [7]

L'image illustre le concept de l'apprentissage non supervisé en apprentissage automatique, mettant en évidence des données brutes non étiquetées en entrée, un algorithme d'apprentissage qui analyse ces données sans connaître le résultat attendu, et des motifs rouges en sortie, représentant des résultats non prédéfinis.

1.3.2.3.L'apprentissage semi-supervisé

L'apprentissage automatique semi-supervisé combine des méthodes supervisées et non supervisées, utile lorsque les données non étiquetées sont nombreuses et l'annotation des données est laborieuse. Contrairement à l'apprentissage supervisé traditionnel qui utilise uniquement des données étiquetées, l'apprentissage semi-supervisé tire parti à la fois des données étiquetées et non étiquetées pour améliorer la précision du modèle [7]. Par exemple, un algorithme non-supervisé de groupement identifie des groupes, puis il attribue une étiquette à chacun des groupes pour ensuite étiqueter tous les membres de chacun de ces groupes [w12]. Les tâches courantes de l'apprentissage semi-supervisé incluent la classification et la régression. Les algorithmes populaires incluent les graphes de propagation de l'étiquette (label propagation), les modèles génératifs et les autoencodeurs variationnels [w11].

² Le clustering, ou regroupement en français, est une technique d'analyse de données utilisée en apprentissage automatique (machine learning) et en statistiques.

1.3.2.4.L'apprentissage par renforcement

Le renforcement par apprentissage est une branche de l'apprentissage automatique qui se concentre sur la manière dont les agents logiciels prennent des actions dans un environnement pour maximiser une récompense cumulative. C'est l'un des trois principaux paradigmes de l'apprentissage automatique, avec l'apprentissage supervisé et non supervisé [7]. Ci-dessous on a Cette image illustre le concept de l'apprentissage par renforcement, où un agent interagit avec un environnement en effectuant des actions et en recevant des récompenses pour apprendre à prendre de meilleures décisions.

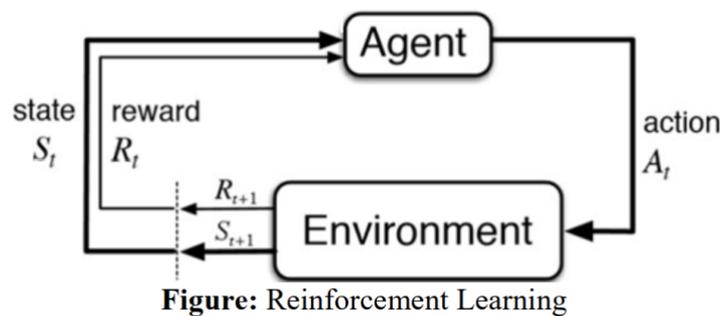


Figure: Reinforcement Learning

Figure 5: L'apprentissage par renforcement [7]

Ce schéma met en évidence le principe de l'apprentissage par renforcement, dans lequel un agent agit au sein d'un environnement en se basant sur les états observés pour maximiser une récompense globale.

1.3.2.5.Apprentissage profond ou Deep Learning

Le Deep Learning, avec son apprentissage en profondeur caractérisé par l'usage de multiples couches dans les réseaux neuronaux, est une méthode informatique puissante. Elle utilise des réseaux de neurones complexes pour comprendre des données telles que des images ou des mots. Cette technique permet aux ordinateurs d'apprendre de manière autonome, en observant de nombreux exemples, un peu comme nous le faisons en tant qu'humains [84]. Grâce à cette approche, les ordinateurs peuvent accomplir diverses tâches, telles que la reconnaissance faciale sur des photos ou la compréhension d'instructions écrites.

Cette technologie peut être appliquée dans différents domaines :

- Dans la vision par ordinateur, elle est utilisée pour la reconnaissance faciale, la détection d'objets, la classification d'images, et la segmentation d'images médicales ;
- En traitement du langage naturel, elle est employée pour la traduction automatique, l'analyse de sentiment, la génération de texte, la reconnaissance vocale, et les chatbots ;



- Dans le domaine de la santé, elle aide au diagnostic médical, à la prédiction de maladies, à l'analyse d'images médicales, et à la découverte de médicaments ;
- Pour l'automobile, elle est utilisée dans la conduite autonome, la détection d'objets sur la route, et la reconnaissance de panneaux de signalisation ;
- Dans le secteur financier, elle sert à faire des prévisions financières, détecter les fraudes, analyser les risques, et effectuer le trading algorithmique ;
- Dans l'industrie, elle est utilisée pour le contrôle de qualité, la maintenance prédictive, et l'optimisation des processus de fabrication ;
- Pour les jeux vidéo, elle est utilisée pour créer des agents de jeu plus intelligents grâce à l'apprentissage par renforcement, ainsi que pour la génération de contenu de jeu.

	Machine Learning	Deep Learning
Organisation des données	Données structurées	Données non structurées
Base de données	Contrôlable	> 1 million de données
Entraînement	Entraînement par l'humain nécessaire	Système d'apprentissage autonome
Algorithme	Algorithme modifiable	Réseau neuronal d'algorithmes
Champ d'application	Actions simples de routine	Tâches complexes

Figure 6: Différence entre le Machine Learning et le Deep Learning [w28].

1.3.3. Synthèse des algorithmes de machine learning

La liste des algorithmes de machine learning est vaste et continue de s'étendre à mesure que de nouvelles recherches sont menées. Le tableau suivant montre une liste de certains des algorithmes de machine learning les plus couramment utilisés, classés en fonction des types d'apprentissage.



<p>Apprentissage Supervisé</p>	<ul style="list-style-type: none"> • Régression linéaire (Linear Regression) ; • Régression logistique (Logistic Regression) ; • Machines à vecteurs de support (Support Vector Machines, SVM) ; • k-Nearest Neighbors (k-NN) • Arbres de décision (Decision Trees) ; • Forêts aléatoires (Random Forests) ; • Réseaux de neurones artificiels (Artificial Neural Networks, ANN) ; • Naive Bayes ; • Réseaux de neurones convolutifs (Convolutional Neural Networks, CNN) ; • Réseaux de neurones récurrents (Recurrent Neural Networks, RNN) ; • Méthodes ensemblistes (Ensemble Methods) comme le Bagging et le Boosting ; • Etc.
<p>Apprentissage Non Supervisé</p>	<ul style="list-style-type: none"> • K-Means ; • Mélange de Gaussiennes (Gaussian Mixture Models, GMM) ; • Analyse en composantes principales (Principal Component Analysis, PCA) ; • Réduction de la dimensionalité (Dimensionality Reduction) ; • Clustering hiérarchique (Hierarchical Clustering) ; • Autoencodeurs ; • Factorisation en matrices non négatives (Non-negative Matrix Factorization, NMF) ; • Cartes auto-organisatrices (Self-Organizing Maps, SOM); • Etc.
<p>Apprentissage par Renforcement</p>	<ul style="list-style-type: none"> • Q-Learning ; • SARSA (State-Action-Reward-State-Action); • Deep Q Network (DQN) ; • Méthodes de gradient de politique (Policy Gradient Methods) ; • Actor-Critic ; • Algorithme génétique pour l'apprentissage par renforcement (Genetic Algorithm for Reinforcement Learning) • Etc.
<p>Apprentissage Semi-Supervisé</p>	<ul style="list-style-type: none"> • Méthodes de transduction (Transductive Methods) • Self-training; • Multi-view learning; • Co-training ; • Etc.

Tableau 1: Tableau contenant les algorithmes de ML les plus utilisées

NB : Le choix d'un algorithme dépend généralement du type de tâche à accomplir, de la nature des données et des contraintes spécifiques du problème.



1.4. Les algorithmes de ML pour la segmentation des clients

Les algorithmes de machine learning pour la segmentation des clients sont des méthodes informatiques qui examinent attentivement les informations sur les clients, les trient en groupes similaires en fonction de leurs caractéristiques communes, comme leurs habitudes d'achat ou leurs préférences, ce qui permet de mieux comprendre leurs comportements et besoins. Ces segments homogènes sont ensuite utilisés pour personnaliser les stratégies marketing, telles que l'envoi d'offres spéciales ou la création de messages ciblés, afin d'optimiser l'engagement et la fidélité des clients.

1.4.1. Les algorithmes de sélections

Ces algorithmes sont des méthodes qui identifient les variables les plus pertinentes dans les données client afin d'améliorer la précision et la pertinence des modèles de segmentation. Parmi les algorithmes de sélection de machine learning couramment utilisés pour la segmentation des clients on peut évoquer :

- **Le clustering K-Means**³ : c'est une technique qui segmente les données en clusters en fonction de leurs similitudes [18]. En effet, elle est souvent utilisée pour regrouper des clients en fonction de leurs comportements d'achat ou de leurs caractéristiques démographiques ;
- **L'Analyse des Composantes Principales (PCA)**⁴ : c'est une technique qui permet de réduire la dimensionnalité des données en identifiant les combinaisons linéaires des variables qui expliquent le plus de variance [19]. Elle est souvent utile pour identifier les caractéristiques les plus importantes afin de segmenter les clients ;
- **L'Analyse Factorielle Discriminante (AFD)**⁵ : c'est une technique qui réduit la dimensionnalité des données tout en maximisant la séparation entre les groupes de données [20]. Cette méthode est bénéfique pour identifier des segments distincts parmi les clients ;
- **Les arbres de décision**⁶ : ce sont des méthodes qui séparent de manière répétitive les données en fonction des attributs les plus significatifs [21]. Cette approche est utile pour classer les clients en fonction de leurs caractéristiques et de leurs comportements ;
- **Les forêts d'arbres de décision**⁷ : également appelées random forests élaborent plusieurs arbres de décision et fusionnent leurs prédictions afin d'améliorer la précision et la fiabilité

³ [sklearn.cluster.KMeans — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html)

⁴ [sklearn.decomposition.PCA — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html)

⁵ [sklearn.discriminant_analysis.LinearDiscriminantAnalysis — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html)

⁶ [sklearn.tree.DecisionTreeClassifier — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html)

⁷ [sklearn.ensemble.RandomForestClassifier — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)



[22]. Cette approche est bénéfique pour segmenter les clients dans des ensembles de données complexes ;

- **L'Analyse de Cluster Hiérarchique**⁸ : c'est une méthode qui regroupe les données en clusters de façon hiérarchique, en mettant en évidence les similitudes entre les individus [23]. Cette approche permet d'identifier des segments de clients à différents niveaux de similarité ;
- **Les Mélanges de Gaussiennes**⁹ : c'est une technique qui modélise les données en les considérant comme provenant d'un mixage de plusieurs distributions gaussiennes [24]. Cette méthode est utile pour identifier des clusters de clients présentant des distributions de caractéristiques différentes.

Ces méthodes peuvent être employées individuellement ou en combinaison, en fonction de la complexité des données et des objectifs spécifiques de segmentation des clients.

1.4.2. Les algorithmes de clustering

Ce sont des techniques qui organisent les clients en groupes similaires en se basant sur leurs caractéristiques communes, ce qui facilite la compréhension des différents segments de clientèle. Parmi les algorithmes de clustering de machine learning utilisés pour la segmentation des clients, on peut citer :

- **Le K-Means**¹⁰ : c'est une méthode largement adoptée pour segmenter les clients. Cette technique divise les données en un nombre préétabli de clusters en minimisant la variance à l'intérieur de chaque cluster [25] ;
- **Le K-Modes**¹¹ : c'est une variante du K-Means conçue spécifiquement pour les données catégorielles. Il s'avère utile lorsque les caractéristiques des clients sont catégoriques plutôt que continues [26] ;
- **Le DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**¹² : c'est un algorithme qui repère les clusters dans des zones de haute densité tout en détectant les valeurs aberrantes. Il s'avère bénéfique lorsque les clusters présentent des formes complexes ou que leurs densités diffèrent [27]. Elle est souvent utilisée dans la segmentation des clients.

⁸ [sklearn.cluster.AgglomerativeClustering — scikit-learn 1.4.1 documentation](#)

⁹ [2.1. Gaussian mixture models — scikit-learn 1.4.1 documentation](#)

¹⁰ [sklearn.cluster.KMeans — scikit-learn 1.4.1 documentation](#)

¹¹ [GitHub - nicodv/kmodes: Python implementations of the k-modes and k-prototypes clustering algorithms, for clustering categorical data](#)

¹² [sklearn.cluster.DBSCAN — scikit-learn 1.4.1 documentation](#)



- **Le Mean Shift¹³** : il identifie les clusters en déplaçant de manière itérative les centres de densité vers les zones de densité maximale dans l'espace des données [28]. Cette méthode est performante pour les ensembles de données de taille moyenne à grande ;
- **Le Clustering Hiérarchique¹⁴** : il crée une structure hiérarchique de clusters en regroupant de manière itérative les points similaires [23]. Cette méthode peut être soit agglomérative (ascendante), soit divisive (descendante) ;
- **Les Modèles de Mélange Gaussien (GMM)¹⁵** : ils supposent que les données proviennent d'un mélange de plusieurs distributions gaussiennes [24]. Cette méthode peut détecter des clusters de formes complexes et est fréquemment utilisée lorsque les clusters suivent une distribution de probabilité gaussienne ;
- **Le Clustering Agglomératif (clustering ascendant)¹⁶** : il commence par traiter chaque point comme un cluster individuel, puis fusionne progressivement les clusters en fonction de leur similarité[29]. Elle permet principalement de regrouper automatiquement les clients en segments homogènes en fonction de leurs comportements et caractéristiques similaires.

Ces méthodes peuvent être appliquées pour segmenter les clients en se basant sur leurs caractéristiques démographiques, comportementales, ou tout autre attribut pertinent. Le choix de l'algorithme est souvent dicté par la nature des données et les objectifs spécifiques de la segmentation.

1.4.3. Les algorithmes de classification

Ce sont des techniques qui classent chaque client dans des groupes spécifiques en se basant sur leurs caractéristiques, ce qui aide à mieux comprendre leurs besoins et à personnaliser les interactions. Parmi les algorithmes de classification, nous avons :

- **Les réseaux de neurones profonds (DNN)** : ce sont des modèles d'apprentissage automatique qui utilisent des architectures complexes inspirées du cerveau humain pour apprendre à classer les données en différentes catégories [30]. Dans le contexte de la segmentation des clients, ces algorithmes sont utilisés pour identifier des groupes de clients distincts en fonction de leurs caractéristiques et de leur comportement ;
- **Les modèles bayésiens** : ils utilisent le théorème de Bayes pour estimer la probabilité qu'un client appartienne à un segment particulier. Ils se basent sur des distributions de probabilité pour modéliser les caractéristiques des clients et les relations entre ces

¹³ [sklearn.cluster.MeanShift — scikit-learn 1.4.1 documentation](#)

¹⁴ [sklearn.cluster.AgglomerativeClustering — scikit-learn 1.4.1 documentation](#)

¹⁵ [2.1. Gaussian mixture models — scikit-learn 1.4.1 documentation](#)

¹⁶ [sklearn.cluster.AgglomerativeClustering — scikit-learn 1.4.1 documentation](#)



caractéristiques et les segments [31]. De plus elles s'adaptent continuellement aux nouvelles données pour améliorer la précision des segments.

- **Le Support Vector Machines (SVM)** : ce sont des algorithmes d'apprentissage automatique supervisé utilisés pour la classification et la régression [32]. Dans le contexte de la segmentation des clients, les SVM sont utilisés pour classer les clients en différents segments en fonction de leurs caractéristiques ;
- **Le random Forest (forêt aléatoire)** : est un algorithme de classification de Machine Learning qui s'appuie sur une multitude d'arbres de décision pour générer des prédictions plus précises et robustes [22];
- **Le K-Nearest Neighbors (KNN)** : c'est un algorithme de classification simple et largement utilisé pour la segmentation des clients. Il fonctionne en se basant sur l'idée que les points de données similaires se regroupent [33] ;
- **Le XGBoost**, ou eXtreme Gradient Boosting, est un algorithme d'apprentissage automatique de type "boosting" utilisé pour la classification et la régression [34]. Il est basé sur les arbres de décision et s'est avéré très performant dans de nombreux domaines, notamment la segmentation des clients.

1.4.4. Les modèles prédictifs

Ce sont des techniques qui se basent sur les données antérieures des clients pour anticiper leur comportement à venir et les classer dans des groupes selon leurs habitudes et leurs préférences. Parmi les algorithmes de modèles prédictifs de machine learning souvent utilisés pour la segmentation des clients on a :

- K-Means Clustering ;
- Mélanges de Gaussiennes (Gaussian Mixture Models) ;
- Analyse des Composantes Principales (Principal Component Analysis - PCA) ;
- Forêts aléatoires (Random Forests) ;
- Machines à Vecteurs de Support (Support Vector Machines - SVM) ;
- Réseaux Neuronaux.

Le choix du modèle dépend de la nature des données et de l'entreprise en ce qui concerne ses objectifs spécifiques.

L'apprentissage continu sur les données clients

L'apprentissage continu consiste à appliquer et à évaluer en permanence les modèles de machine learning en prenant en compte les nouvelles données clients, ce qui assure une



segmentation toujours à jour et plus précise [9]. Parmi les principaux algorithmes d'apprentissage continu de machine learning utilisés pour la segmentation des clients on peut citer :

- **L'apprentissage incrémental avec K-Means**¹⁷ implique la mise à jour progressive d'un modèle de clustering tel que K-Means à mesure que de nouvelles données sur les clients sont disponibles [35]. Plutôt que de réentraîner le modèle à partir de zéro chaque fois qu'une nouvelle donnée est ajoutée, le modèle existant peut être mis à jour itérativement pour intégrer les nouvelles données. Comme exemple on a :
 - **K-Means Streaming** : c'est un algorithme de machine learning en ligne utilisé pour la segmentation des clients en temps réel ;
 - **Mini-batch K-Means**: c'est une variante de l'algorithme K-Means classique qui s'adapte à l'apprentissage continu et à la segmentation des clients en temps réel ;
 - **Etc.** ;
- **L'apprentissage en ligne avec les réseaux de neurones** : implique l'entraînement continu de ces modèles à mesure que de nouvelles données sont reçues. Grâce à des méthodes d'apprentissage en ligne, le modèle peut être mis à jour à chaque nouvelle observation, ce qui lui permet de s'adapter rapidement aux évolutions dans les comportements des clients [36] ;
- **L'adaptation des modèles de classification pour l'évolution des segments** implique la modification de modèles tels que les arbres de décision ou les SVM afin de tenir compte des changements dans les segments de clients. En utilisant des techniques d'adaptation de modèle [37], le modèle peut être ajusté pour refléter les nouveaux segments émergents ou les modifications dans les préférences des clients ;
- **La mise à jour en temps réel des modèles de prédiction, tels que les forêts aléatoires ou les modèles de régression logistique** : il consiste à ajuster ces modèles à mesure que de nouvelles données sont collectées [38]. Cette approche permet aux modèles de s'adapter rapidement aux évolutions dans les comportements des clients, assurant ainsi des prédictions précises et actualisées ;
- **L'intégration du retour utilisateur dans l'apprentissage** consiste pour les systèmes de recommandation et de personnalisation à actualiser de manière continue les modèles de segmentation des clients en se basant sur ce retour[38]. En recourant à des méthodes telles que le filtrage collaboratif ou le reinforcement learning, les modèles peuvent évoluer pour s'ajuster aux préférences individuelles des clients au fil du temps ;

¹⁷ [sklearn.cluster.MiniBatchKMeans — scikit-learn 1.4.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MiniBatchKMeans.html)



- **Birch Clustering** : c'est un algorithme de clustering hiérarchique utilisé pour la segmentation de données avec des dimensions élevées [39]. Elle le fait de manière hiérarchique, progressive et permet de segmenter rapidement et efficacement même avec de grandes quantités de données. ;
- **HDBSCAN** : c'est un algorithme d'apprentissage automatique de type "clustering" utilisé pour segmenter des données en groupes naturels. Dans la segmentation des clients, il permet de répartir les clients en identifiant des groupes de densité variable, permettant de trouver de différentes tailles et formes sans avoir à spécifier le nombre de segments à l'avance [40] ;
- **OPTICS** (Ordering Points To Identify the Clustering Structure) : est un algorithme d'apprentissage automatique de type "clustering" utilisé pour la segmentation des clients. Il répartit les clients en analysant la densité des données pour identifier des groupes de différentes tailles et formes, sans avoir besoin de prédéfinir le nombre de segments [27] ;

Ces exemples démontrent comment l'apprentissage continu peut servir à améliorer la segmentation des clients en se ajustant aux évolutions dans leurs comportements et leurs préférences au fil du temps.



Chapitre 2 : Le Machine Learning et la Segmentation des clients



2.1. Les apports du machine learning dans la segmentation des clients

Le machine learning joue un rôle primordial dans la segmentation des clients. En effet, il permet d'analyser des volumes importants de données et de regrouper les clients aux caractéristiques similaires. Grâce au machine learning, la segmentation des clients devient plus précise et facile, facilitant également la communication. De plus, l'apport principal du machine learning est de mieux comprendre les clients pour répondre à leurs besoins et attentes, permettant ainsi de prendre les meilleures décisions stratégiques [9]. Voici les principaux apports du machine learning dans la segmentation des clients :

- Analyse de grands volumes de données : Le machine learning permet de traiter et d'analyser des quantités massives de données rapidement et efficacement ;
- Segmentation précise : Il permet de créer des segments de clients plus précis basés sur des caractéristiques et des comportements spécifiques ;
- Automatisation de la segmentation : Le processus de segmentation peut être automatisé, réduisant ainsi le temps et les efforts nécessaires pour classer les clients ;
- Amélioration de la communication : La segmentation basée sur le machine learning permet de personnaliser les communications et les offres en fonction des segments de clients ;
- Compréhension approfondie des clients : Il aide à mieux comprendre les comportements, les besoins et les préférences des clients ;
- Prédiction des comportements futurs : Le machine learning peut prédire les comportements futurs des clients, ce qui aide à anticiper les tendances et les besoins ;
- Optimisation des stratégies marketing : En fournissant des insights détaillés sur les segments de clients, il permet d'optimiser les campagnes marketing et d'allouer les ressources de manière plus efficace ;
- Réduction des coûts : En automatisant la segmentation et en ciblant plus précisément les clients, il contribue à réduire les coûts marketing ;
- Augmentation de la satisfaction client : En personnalisant les offres et les communications, il aide à mieux répondre aux attentes des clients, améliorant ainsi leur satisfaction ;
- Détection de nouveaux segments : Le machine learning peut identifier des segments de clients qui n'étaient pas évidents auparavant, ouvrant de nouvelles opportunités de marché ;
- Amélioration de la fidélisation : En comprenant mieux les besoins et les comportements des clients, il aide à développer des stratégies de fidélisation plus efficaces ;
- Réduction du churn : En prédisant quels clients sont susceptibles de partir, il permet de mettre en place des actions préventives pour réduire le taux de désabonnement.



2.1.1. Amélioration de la précision dans les segments

L'amélioration de la précision des segments consiste à affiner les groupes de clients en utilisant des techniques avancées de machine learning et d'analyse de données. Cela garantit que chaque segment regroupe des individus aux caractéristiques et comportements très similaires, rendant les actions marketing et les stratégies de communication plus ciblées et efficaces. Par conséquent, les entreprises peuvent mieux comprendre le comportement et les préférences des clients, adaptant ainsi leurs stratégies marketing pour maximiser leurs résultats [w14]. De plus, il est essentiel d'analyser et d'affiner régulièrement ces segments en fonction des évolutions du marché et des besoins des clients. Cette approche renforce non seulement la gestion stratégique de l'entreprise, mais aussi sa compétitivité sur le marché [10]. Par exemple, une boutique en ligne comme Jumia peut utiliser le machine learning pour repérer les clients qui achètent souvent des vêtements de marque, mais seulement à certaines saisons. En comprenant cela, la boutique peut leur envoyer des offres spéciales au bon moment, ce qui les encourage à acheter d'avantage [w14].

2.1.2. Segmentation dynamique

La segmentation dynamique des clients implique de diviser la base de clients en groupes en constante évolution, prenant en compte les changements dans leurs comportements, préférences et besoins au fil du temps. Contrairement aux méthodes traditionnelles qui reposent sur des critères statiques, les modèles d'apprentissage automatique offrent l'avantage de la segmentation dynamique [11]. Ainsi, les segments peuvent être rapidement ajustés pour refléter les évolutions du comportement des clients, permettant aux entreprises de proposer des expériences personnalisées en temps réel [w14]. Cette capacité à identifier et à répondre rapidement aux changements dans les préférences des clients peut aider les entreprises comme les agences de voyage à rester compétitives et à fidéliser leur clientèle en offrant des expériences de voyage adaptées et personnalisées. De plus, le machine learning facilite la répartition des clients selon des critères spécifiques, améliorant ainsi la précision de la segmentation. Un exemple pertinent est celui du secteur touristique, où une agence de voyage en ligne comme Sénégal Original¹⁸ pourrait tirer parti des technologies de suivi et d'analyse pour détecter les évolutions dans les préférences de voyage d'un client. Par exemple, si un client manifeste un intérêt croissant pour des destinations exotiques après avoir consulté plusieurs offres de voyages balnéaires traditionnels, l'agence pourrait ajuster ses recommandations en conséquence. Ainsi, elle serait en mesure de proposer des

¹⁸ [Voyage Sénégal - Agence de voyage locale - Sénégal Original \(senegal-original.com\)](http://senegal-original.com)



forfaits de voyage personnalisés pour des destinations plus aventureuses ou culturellement riches, répondant de manière proactive aux nouvelles aspirations de voyage du client.

2.1.3. Analyse prédictive

L'analyse prédictive aide les responsables marketing à anticiper les besoins des clients et à optimiser leurs campagnes. En utilisant des algorithmes de machine learning, elle permet de personnaliser les offres et d'améliorer l'efficacité des stratégies marketing [12]. Grâce aux algorithmes d'apprentissage automatique, il est possible de prédire avec précision les comportements et les préférences futurs des clients en analysant de vastes ensembles de données [w14]. Ainsi, les entreprises peuvent adapter leurs stratégies marketing, proposer des produits adaptés et offrir une expérience client plus personnalisée, tout en comprenant mieux les besoins individuels des clients. Par exemple, en analysant les habitudes de dépenses et les transactions passées des utilisateurs, des plateformes telles que Wizall Money, Orange Money, Tigo Cash et K Pay, qui sont des services financiers numériques sénégalais, peuvent fournir des recommandations personnalisées sur les produits et services financiers susceptibles de les intéresser. De plus, ces plateformes peuvent offrir des conseils pour gérer plus efficacement leur argent, améliorant ainsi la satisfaction et la fidélité des clients. L'analyse prédictive aide les responsables marketing à anticiper les besoins des clients et à optimiser leurs campagnes. En utilisant des algorithmes de machine learning, elle permet de personnaliser les offres et d'améliorer l'efficacité des stratégies marketing.

2.1.4. La personnalisation des offres

La personnalisation des offres grâce au machine learning consiste à utiliser des algorithmes et des techniques d'analyse de données pour créer des recommandations sur mesure, adaptées aux besoins, préférences et comportements individuels des clients. Cette approche est essentielle pour segmenter les clients de manière efficace. En effet, en analysant les données démographiques, les historiques d'achats et les interactions en ligne, les algorithmes de machine learning permettent aux entreprises de développer des offres sur mesure pour chaque segment [13]. Par conséquent, ces offres personnalisées améliorent l'expérience utilisateur, favorisent la fidélisation et augmentent les revenus. En somme, le machine learning est un outil puissant pour optimiser le marketing et répondre efficacement aux attentes des clients.

2.2. Les enjeux

2.2.1. La Tarification dynamique

La tarification dynamique consiste à ajuster les prix des produits ou services en temps réel selon la demande du marché, la période, l'accessibilité des produits et parfois le comportement



des clients, permettant ainsi aux entreprises de maximiser leurs rendements. En effet, Les stratégies de tarification peuvent être optimisées en fonction des préférences individuelles des clients et des conditions du marché grâce aux algorithmes d'apprentissage automatique [14]. Par exemple, les compagnies aériennes sénégalaises pourraient utiliser ces algorithmes pour fixer des prix personnalisés pour les billets, en tenant compte de facteurs comme la demande, les profils des clients et les données historiques sur les prix.

2.2.2. La Publicité ciblée

La publicité ciblée adapte les annonces en fonction des données démographiques, comportementales et contextuelles des clients pour répondre précisément à leurs intérêts et besoins, ce qui accroît l'efficacité des campagnes publicitaires. En effet, les algorithmes d'apprentissage automatique jouent un rôle crucial dans cette optimisation, en analysant les préférences individuelles et les comportements pour diffuser des publicités personnalisées qui vont avec le choix de chaque consommateur. Par exemple, des plateformes sénégalaises telles que Jumia Sénégal ou CoinAfrique utilisent ces algorithmes avancés pour ajuster leurs annonces selon les intérêts spécifiques des utilisateurs, basés sur leurs interactions et préférences sur le site.

2.2.3. La Recommandations de produits

Les algorithmes d'apprentissage automatique jouent un rôle essentiel dans l'amélioration de l'expérience d'achat des clients en recommandant des produits personnalisés, en analysant les préférences et les comportements passés des clients. En effet, la recommandation des produits s'agit de proposer des produits ou services pertinents en analysant les préférences et les comportements passés des clients [15]. En exploitant les données client et en identifiant les schémas d'achat, les entreprises peuvent suggérer des produits pertinents qui correspondent aux préférences individuelles [w14]. Par exemple, des enseignes comme Auchan Sénégal et Decathlon pourraient utiliser ces technologies pour proposer des recommandations personnalisées, augmentant ainsi les ventes et la satisfaction client.

2.2.4. Les Campagnes marketing personnalisées

Les campagnes marketing personnalisées, lorsqu'appliquées à la segmentation des clients par machine learning, consistent à utiliser des algorithmes avancés pour analyser les données client et identifier des segments ayant des comportements et des préférences similaires. En comprenant ces segments, les entreprises peuvent adapter leurs messages et leurs offres de manière précise, améliorant ainsi l'efficacité de leurs stratégies marketing en augmentant l'engagement et la satisfaction client tout en optimisant les résultats commerciaux. Par exemple, un détaillant comme



tong tong¹⁹ pourrait adapter ses campagnes par e-mail en fonction de l'historique de navigation et d'achat de chaque client, ce qui a prouvé son efficacité pour augmenter l'engagement, la fidélité et les ventes.

2.2.5. La Fidélisation améliorée des clients

Dans le contexte de la segmentation par apprentissage automatique, la fidélisation améliorée des clients implique l'utilisation d'algorithmes avancés pour identifier les clients à haut risque de résiliation (churn) et ceux ayant un fort potentiel de fidélisation. En analysant les données comportementales et transactionnelles, les entreprises peuvent mettre en place des stratégies personnalisées telles que des programmes de fidélité ciblés et des offres exclusives [16]. L'objectif est de renforcer les relations avec leurs clients les plus précieux et de réduire le taux d'attrition [w14]. Par exemple, en appliquant une campagne de fidélisation améliorée, comprenant des offres spéciales ou des réductions, les entreprises de télécommunications comme Sonatel (Orange), Expresso Sénégal et Free Sénégal peuvent détecter une baisse significative de l'activité chez un client. Cette approche proactive a le potentiel de réduire considérablement le taux de résiliation des abonnements et d'améliorer la satisfaction globale des clients.

2.2.6. La Prise de décision en temps réel

La segmentation client en temps réel basée sur l'apprentissage automatique permet aux entreprises de prendre des décisions agiles en analysant en continu les données des clients. Grâce à l'utilisation d'algorithmes pour analyser instantanément les données et prendre des décisions immédiates en fonction des résultats obtenus, les entreprises peuvent ajuster rapidement leurs stratégies marketing, promotions et interactions avec les clients en fonction des comportements actuels et des conditions du marché [17]. Par exemple, une plateforme de streaming comme Wido ou IrokoTV pourrait rapidement adapter ses stratégies pour répondre aux préférences des utilisateurs, notamment ceux intéressés par les séries sénégalaises et africaines. Elle pourrait lancer des campagnes publicitaires ciblées pour promouvoir ces séries, démontrant ainsi sa réactivité et sa capacité à s'adapter aux demandes du marché. Cette flexibilité permettrait aux entreprises de rester compétitives et de saisir les opportunités qui se présentent.

2.3. Les contraintes liées à la mise en œuvre

2.3.1. Accès et Disponibilité des données

L'accès et la disponibilité des données sont des éléments essentiels pour la segmentation des clients par l'apprentissage automatique (ML). Ils déterminent la capacité à obtenir et à utiliser

¹⁹ [Tong Tong](#)



les données nécessaires pour entraîner et déployer efficacement des modèles d'apprentissage automatique. Ces données utilisées pour segmenter la clientèle proviennent souvent de sources internes (données transactionnelles, CRM, données démographiques, d'engagement) et externes (données sociales, d'études de marché, d'achats en ligne, météorologiques, économiques), offrant ainsi une diversité et une richesse considérables. Par exemple, on peut citer les transactions d'achat, les interactions CRM, ainsi que les données démographiques et sociales.

Malgré les nombreux avantages de la segmentation client par apprentissage automatique, l'accès et la disponibilité des données nécessaires à sa mise en œuvre rencontrent plusieurs défis. Parmi ceux-ci figurent la limitation d'accès, l'existence de données incomplètes, inexacts ou erronées, ainsi que des problèmes d'incompatibilité de format de données avec les outils de ML. De plus, il peut y avoir un manque de qualifications et de ressources adéquates pour gérer ces données. En surmontant ces défis, les entreprises peuvent exploiter pleinement le potentiel de l'apprentissage automatique pour segmenter la clientèle avec précision, répondre à leurs besoins, prédire et personnaliser les offres et les demandes, tout en respectant les normes de confidentialité et de sécurité des données établies.

2.3.2. Qualité des données

La qualité des données est une mesure de l'état des données, évaluée selon des critères tels que la précision, l'exhaustivité, l'homogénéité, la fiabilité et l'actualité, permettant aux organisations de repérer et de corriger les erreurs potentielles tout en déterminant si les données sont appropriées pour leurs besoins [w33]. Avant d'être utilisées par les technologies d'intelligence artificielle, notamment l'apprentissage automatique, les données primaires doivent subir un processus rigoureux de nettoyage et de préparation. Cela implique l'élimination des doublons, la correction des erreurs, la vérification du format des données et l'assurance que la quantité de données est suffisante, ce qui est crucial pour la précision des modèles de segmentation client.

2.3.3. Manque d'expertises en ML et en data science

Avec l'avènement du marketing digital et la prolifération des données clients, la segmentation client par machine learning ouvre de nouvelles perspectives pour les entreprises, leur permettant de segmenter leurs clients de manière plus précise et efficace grâce aux technologies d'intelligence artificielle. Cependant, de nombreuses entreprises se heurtent à un défi majeur : le manque d'expertise en apprentissage automatique (ML) et en science des données. Ces domaines spécialisés exigent des compétences pointues, notamment la collecte, le nettoyage et l'analyse des données, ainsi que le développement et le déploiement de modèles de ML, indispensables pour traduire les résultats en actions marketing concrètes.



L'absence de connaissances spécialisées en ML et en science des données peut entraver la segmentation client de plusieurs façons, telles que la difficulté à identifier les segments pertinents, l'inefficacité des modèles de ML et la compréhension limitée des résultats obtenus. Pour pallier ce manque, les entreprises peuvent envisager diverses solutions comme le recrutement de data scientists, l'externalisation de la segmentation client, ou encore la formation des équipes marketing aux concepts fondamentaux du ML. Investir dans le développement d'expertise en ML et en science des données est crucial pour les entreprises souhaitant tirer parti de la segmentation client afin d'améliorer leurs performances marketing.

2.3.4. Aspects légaux de collecte et d'utilisations des données personnelles

Les entreprises doivent se conformer strictement aux réglementations en vigueur concernant les données personnelles des clients afin d'éviter des sanctions et de protéger leur vie privée. Parmi les aspects juridiques essentiels, le consentement éclairé des clients est crucial avant toute collecte et utilisation des données, en fournissant des informations transparentes sur les types de données collectées, leurs finalités, les partages éventuels, ainsi que les options de contrôle pour les clients. De plus, la protection des données requiert la mise en place de mesures de sécurité adéquates telles que le cryptage et le contrôle d'accès. Les entreprises doivent également veiller à ce que leurs algorithmes de ML n'induisent aucune forme de discrimination, et garantir la transparence quant à leur utilisation pour la segmentation des clients, incluant les critères utilisés et les implications potentielles pour les clients [83]. Enfin, elles doivent respecter le droit des clients à accéder à leurs données et à les faire supprimer, tout en intégrant des considérations éthiques pour assurer des pratiques équitables et transparentes tout en respectant la vie privée des clients.



Chapitre 3 : Les outils de machine learning prêts à l'emploi pour la segmentation des clients.



3.1. Les outils de machine learning prêts à l'emploi pour la segmentation des clients

3.1.1. Outils d'analyse de données avec le ML

L'analyse de données consiste à explorer, traiter et modéliser les données pour identifier des groupes de clients similaires, permettant ainsi une segmentation précise et personnalisée. Comme exemples d'outils d'analyse de données de machine learning utilisés pour la segmentation des clients on peut citer :

- **Google Analytics**²⁰ : c'est un outil gratuit d'analyse d'audience web proposé par Google. Il nous permet de mesurer le trafic sur le site web et d'obtenir des informations précieuses sur les visiteurs. permet d'identifier les segments de clients en fonction de leur navigation sur le site web, analyser les sources de trafic, mesurer l'impact de vos campagnes marketing [41]. Ses principales fonctionnalités sont l'analyse du trafic web, segmentation des audiences, suivi des conversions, analyse des comportements des utilisateurs ;
- **Adobe Analytics**²¹: c'est une solution d'analyse web et mobile complète qui offre des fonctionnalités de segmentation avancées basées sur le Machine Learning (ML) proposée par Adobe qui offre un éventail plus large de fonctionnalités pour les entreprises qui ont besoin d'une vue plus complète de leur marketing digital et de l'expérience client [42]. Il permet d'identifier les segments de clients en fonction de leurs interactions avec vos différents canaux marketing, analyser les données CRM, créer des profils clients enrichis. Ses principales fonctionnalités sont l'analyse des données web et mobiles, segmentation des audiences, analyse des comportements des utilisateurs, création de rapports personnalisés ;
- **Segment**²²: c'est une plateforme de données client (CDP) qui permet de collecter, d'unifier et d'activer les données clients à partir de différentes sources. Il permet de créer des segments de clients en fonction de leurs données démographiques, comportementales et psychographiques, envoyer des emails personnalisés, diffuser des publicités ciblées [43]. Ses principales fonctionnalités sont la collecte de données clients, segmentation des audiences, création de profils clients, activation des données clients ;
- **Exponea**²³: c'est une plateforme de marketing automation²⁴ et d'analyse de données qui utilise le Machine Learning (ML) pour vous aider à mieux comprendre vos clients et à

²⁰ [Analytics \(google.com\)](https://analytics.google.com/)

²¹ [Adobe Analytics | Web Analytics pour une meilleure aide à la décision](https://adobe.com/learn/topics/web-analytics)

²² <https://segment.com/>

²³ <https://exponea.com/>

²⁴ C'est le fait d'automatiser les tâches marketing par le biais d'outils technologiques pour la personnalisation de la clientèle.



segmenter votre audience. Il permet d'identifier les segments de clients à fort potentiel, proposer des produits et services personnalisés, automatiser les campagnes marketing [44]. Ses principales fonctionnalités sont l'analyse des données clients, segmentation des audiences, création de recommandations personnalisées, automatisation du marketing ;

- **Klaviyo**²⁵: c'est une plateforme de marketing automation e-commerce qui propose des outils d'analyse de données basés sur le Machine Learning (ML) pour la segmentation des clients. Il permet d'envoyer des emails personnalisés aux segments de clients identifiés, créer des scénarios automatisés en fonction du comportement des clients, analyser les résultats de vos campagnes e-mailing [44]. Ses principales fonctionnalités sont le marketing automation par email, segmentation des audiences, création de workflows personnalisés, analyse des performances des campagnes ;
- **IBM SPSS Modeler**²⁶ : est un outil d'analyse de données avec ML qui peut être utilisé pour la segmentation de la clientèle [45]. Il dispose de plusieurs fonctionnalités dont préparer les données, le choix d'un algorithme de segmentation, entraîner et évaluer le modèle de segmentation, déployer le modèle de segmentation ;
- **RapidMiner**²⁷: c'est une plateforme logicielle de science des données et d'apprentissage automatique qui propose des outils puissants pour la segmentation des clients. Ses principales fonctionnalités sont la Préparation des données, Exploration des données, Sélection de variables, Modélisation de segmentation, Évaluation et validation des modèles, Interprétation des segments, Déploiement des modèles, etc. [48] ;
- **WEKA**²⁸(Waikato Environment for Knowledge Analysis) : c'est un outil open-source largement utilisé pour l'exploration de données et l'apprentissage automatique [49]. Elle offre beaucoup d'outils et d'algorithmes pour la segmentation des clients comme : les algorithmes de clustering, les outils de prétraitement des données, les outils d'évaluation du modèle et la visualisation des données ;
- **IBM Watson Studio**²⁹ : c'est une plateforme de cloud computing qui offre un large éventail d'outils et de services pour l'analyse de données, l'apprentissage automatique et l'intelligence artificielle pour la segmentation des clients[50]. Cette plateforme puissante permet aux entreprises de Collecter et intégrer de données, de nettoyer et préparer des

²⁵ <https://www.klaviyo.com/fr/>

²⁶ [Modélisateur IBM SPSS](#)

²⁸ [The Data Platform for Cloud & AI | WEKA - WEKA](#)

²⁹ [IBM Watson Studio](#)



données, d'explorer et d'analyser les données, d'utiliser des techniques d'apprentissage automatique, de créer des segments de les déployer et de les activer.

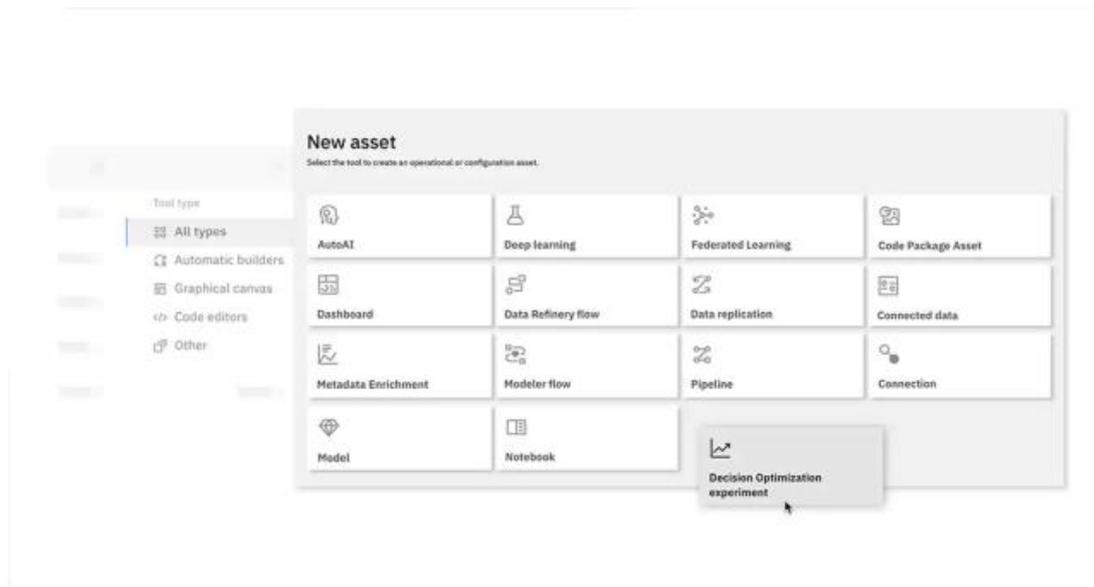


Figure 7 option de segmentation dans ibm Watson studio [w34]³⁰

- **Microsoft Azure Machine Learning**³¹ : c'est une plateforme cloud qui vous permet de construire et déployer des modèles de machine learning pour diverses tâches, dont la segmentation de la clientèle afin de mieux comprendre les clients, de cibler les campagnes marketing et d'améliorer les produits et services [51]. Ses principales fonctionnalités sont la préparation des données, choix de l'algorithme de segmentation, l'entraînement du modèle, l'évaluation du modèle, le déploiement du modèle.

Ces solutions fournissent un large éventail de possibilités pour segmenter les clients, allant des méthodes traditionnelles de machine learning aux techniques avancées de deep learning. Le choix de l'outil dépend généralement des besoins spécifiques de l'entreprise et du niveau de compétence de l'utilisateur en analyse de données.

3.1.2. Outils d'analyse des sentiments avec le ML

L'analyse des sentiments consiste à extraire et à interpréter les sentiments des clients à partir de diverses données non structurées, telles que les avis en ligne ou les commentaires sur les réseaux sociaux, afin de mieux comprendre les préférences et les comportements des clients et d'adapter les stratégies marketing en conséquence. Pour la segmentation des clients en fonction de leurs sentiments, on peut utiliser des outils d'analyse de sentiments de machine learning comme :

³¹[Azure Machine Learning - ML as a Service | Microsoft Azure](#)



- **Google Cloud Natural Language API³²**: c'est un outil puissant qui peut être utilisé pour la segmentation client en analysant le sentiment et la tonalité des données textuelles de vos clients [69]. Avec celui-ci on peut collecter des données textuelles, analyser des sentiments, identifier des thèmes et des entités, segmenter des clients, et enfin faire le ciblage et la personnalisation ;
- **IBM Watson Tone Analyzer³³**: est un outil d'analyse sentimentale qui peut être utilisé pour analyser le ton et les émotions d'un texte. Cela peut être utile pour la segmentation de la clientèle car le ton et les émotions peuvent refléter les attitudes et les besoins des clients. Avec celui-ci on peut analyser le sentiment, détecter des émotions, faire de l'extraction d'aspects³⁴[70] ;
- **Amazon Comprehend³⁵**: c'est un service d'apprentissage automatique (ML) qui offre des fonctionnalités d'analyse du langage naturel (NLP) pour traiter et extraire des informations à partir de textes non structurés. L'une des fonctionnalités de ce service est l'analyse sentimentale, qui permet de déterminer le sentiment global d'un texte (positif, négatif, neutre ou mitigé) [71]. celui-ci peut être utilisé pour la segmentation des clients afin d'identifier les clients insatisfaits, de détecter les changements d'opinion, de segmenter les clients par profil émotionnel, et de pouvoir cibler les campagnes marketing ;
- **Microsoft Azure Text Analytics API³⁶**: c'est un outil précieux pour la segmentation des clients. Il vous permet d'extraire des informations sur les sentiments et les opinions exprimés dans les commentaires des clients, ce qui peut vous aider à les regrouper en segments distincts en fonction de leur attitude envers votre produit ou service. On peut l'utiliser afin d'identifier les clients insatisfaits, de pouvoir segmenter les clients par niveau de satisfaction, de cibler les segments clients avec des offres personnalisées [72] ;
- **Symantic Opinion Mining Suite³⁷**: est un outil d'analyse sentimentale qui peut être utilisé pour la segmentation des clients. Il permet d'extraire des opinions et des sentiments à partir de textes non structurés, tels que les avis clients, les publications sur les réseaux sociaux et les emails [73]. Ces informations peuvent ensuite être utilisées pour segmenter les clients en fonction de leur attitude envers votre marque, vos produits ou vos services. Avec celui-

³² [Cloud Natural Language | Google Cloud | Google Cloud](#)

³³ [Compréhension du langage naturel IBM Watson](#)

³⁴ C'est le processus d'identification et d'extraction des points clés ou des caractéristiques importantes d'un texte pour en faciliter l'analyse et la compréhension.

³⁵ [Traitement du langage naturel – Amazon Comprehend – Amazon Web Services](#)

³⁶ [Langage Azure AI | Microsoft Azure](#)

³⁷ [Exploration d'opinions avec l'analyse de sujets alimentée par l'IA | Symanto](#)



ci on peut faire l'extraction d'opinions et de sentiments, l'identification des thèmes et des sujets, l'attribution d'une note d'opinion et enfin la visualisation des données ;

- **VADER (Valence Aware Dictionary and sEntiment Reasoner)**³⁸ : c'est un outil open-source qui permet d'analyser le sentiment d'un texte, c'est-à-dire de déterminer s'il est positif, négatif ou neutre. il peut être utilisé pour la segmentation des clients en analysant le sentiment des commentaires des clients, des publications sur les réseaux sociaux et d'autres formes de données textuelles [74]. Cela peut vous aider à identifier les clients satisfaits, insatisfaits et neutres. Avec celui-ci on peut collecter les données textuelles des clients, analyser le sentiment de chaque texte, segmenter les clients en fonction de leur score de sentiment, analyser chaque segment pour identifier les caractéristiques et les besoins des clients. ;
- **TextBlob**³⁹ : c'est une bibliothèque Python open-source qui offre des fonctionnalités d'analyse de texte, y compris l'analyse sentimentale qui peut être utiliser afin de segmenter les clients en fonction de leur perception de la marque, du produit ou du service [75]. On peut l'utiliser principalement pour collecter les données textuelles des clients, analyser le sentiment de chaque texte, segmenter les clients en fonction de leur score de sentiment, analyser les segments de clients pour identifier les tendances et les points d'amélioration ;
- **IBM Watson Natural Language Understanding (NLU)**⁴⁰ : c'est un service cloud qui offre des fonctionnalités d'analyse du langage naturel, y compris l'analyse sentimentale [76]. Cette capacité peut être utilisée pour la segmentation de la clientèle en analysant le ton et le sentiment des commentaires des clients, des publications sur les réseaux sociaux et d'autres sources de données textuelles. Celui-ci permet de faire l'extraction des émotions la détection des intentions et l'analyse des thèmes.

Ces bibliothèques et outils offrent une gamme variée d'options pour l'analyse de sentiments dans le texte, ce qui peut être précieux pour segmenter les clients en fonction de leurs émotions, attitudes et opinions exprimées. Le choix de l'outil dépend souvent des performances recherchées, de la langue des textes à analyser, ainsi que de l'intégration avec d'autres systèmes existants.

³⁸ [GitHub - cjhutto/vaderSentiment : Analyse des sentiments VADER. VADER \(Valence Aware Dictionary and sEntiment Reasoner\) est un outil d'analyse des sentiments basé sur un lexique et des règles qui est spécifiquement adapté aux sentiments exprimés dans les médias sociaux et fonctionne bien sur des textes d'autres domaines.](#)

³⁹ [TextBlob : Traitement de texte simplifié — Documentation de TextBlob 0.18.0.post0](#)

⁴⁰ [Compréhension du langage naturel IBM Watson](#)



Chapitre 4 : Etat de l'art et positionnement



4.1. Etat de l'art

Pour obtenir une vue approfondie du domaine, il est essentiel d'analyser la littérature existante. Parmi les études clés qui ont émergé, la segmentation automatique des clients s'est imposée comme une thématique cruciale dans le domaine du marketing et de l'analyse des données, notamment avec l'essor des big data. Plusieurs recherches ont contribué à éclairer cette problématique, en explorant les méthodes de segmentation de la clientèle pour le ciblage personnalisé dans le contexte du commerce électronique [86]. Ces travaux offrent une revue exhaustive des approches de segmentation, tout en soulignant les défis et les opportunités offerts par les techniques d'apprentissage automatique. En outre, leur analyse met en lumière la nécessité d'un cadre méthodologique robuste pour mieux exploiter les données massives et créer des segments de clients pertinents.

Une autre étude se concentre sur l'application de l'apprentissage non supervisé pour la segmentation des clients dans le secteur de l'e-commerce [85]. Les auteurs ont utilisé deux algorithmes principaux, K-means et DBSCAN, pour segmenter les clients en fonction de leurs comportements d'achat. Cette recherche souligne l'importance de choisir le bon algorithme en fonction de la nature des données. En effet, la segmentation des clients, essentielle pour le marketing moderne, a bénéficié des avancées du machine learning, permettant une personnalisation accrue et une meilleure compréhension des comportements d'achat [85].

Cette section explore donc les méthodes traditionnelles de segmentation, les améliorations récentes apportées par le machine learning, ainsi que les défis et perspectives pour l'avenir. Toutefois, il convient de noter que ces méthodes traditionnelles incluent des techniques statistiques telles que le K-means [85], qui reposent souvent sur des analyses basées sur des caractéristiques démographiques et comportementales. Bien qu'elles offrent une base solide, elles sont limitées par leur rigidité et leur incapacité à capturer des dynamiques complexes.

4.1.1. Avancées Récentes avec le Machine Learning

L'introduction du machine learning a transformé la segmentation des clients. Certains auteurs [87] ont montré que les techniques de clustering améliorées, telles que le clustering hiérarchique et les modèles mixtes, peuvent fournir une segmentation plus fine et précise. D'autres chercheurs [88] estiment que l'intelligence artificielle permet aux entreprises de progresser en améliorant à la fois les conditions de travail des employés, grâce à l'automatisation et à la personnalisation, et en optimisant l'expérience client, en rendant les interactions plus fluides et efficaces.



4.1.2. Limitations des Outils Actuels

Bien que des avancées significatives aient été réalisées, plusieurs défis persistent. Certains auteurs [89] notent que les outils actuels de machine learning sont souvent trop complexes et coûteux pour les petites et moyennes entreprises (PME). Cette complexité entrave leur adoption et leur efficacité dans divers contextes commerciaux, soulignant ainsi la nécessité de développer des solutions plus accessibles.

4.1.3. Adaptabilité des Modèles

L'adaptabilité des modèles à différents contextes commerciaux représente une lacune majeure. Certains auteurs [90] ont observé que de nombreux modèles sont optimisés pour des ensembles de données spécifiques, ce qui limite leur applicabilité générale. Cette problématique met en évidence la nécessité d'une recherche approfondie pour développer des modèles flexibles capables de s'ajuster à divers contextes commerciaux. De plus, cette étude [90] examine spécifiquement comment les techniques de segmentation, y compris le clustering, peuvent être appliquées pour analyser et segmenter les bases de données clients dans le cadre d'initiatives marketing et commerciales.

4.1.4. Segmentation Dynamique

Les approches actuelles se concentrent principalement sur des segmentations statiques. Certaines études [92] explorent comment les techniques d'apprentissage automatique peuvent être utilisées pour réaliser une segmentation dynamique des clients, capable de s'adapter aux besoins évolutifs en temps réel. De plus, ces recherches montrent que les modèles de segmentation basés sur l'apprentissage automatique surpassent de manière significative les méthodes traditionnelles en termes de précision et de réactivité face aux changements de comportement des clients. D'autres études [91] soulignent que les données générées par les utilisateurs sur les plateformes sociales, combinées aux techniques analytiques du big data, offrent aux entreprises des insights précieux pour personnaliser leurs offres, anticiper les besoins des clients et améliorer l'engagement. Ces auteurs mettent également en lumière l'importance d'une approche centrée sur les données pour optimiser la performance marketing et réussir dans le contexte de la quatrième révolution industrielle, caractérisée par l'essor de l'intelligence artificielle, de l'automatisation et de l'Internet des objets. Certains d'entre eux [91] vont même jusqu'à proposer des recommandations aux entreprises souhaitant tirer parti de ces technologies pour rester compétitives.



4.1.5. Perspectives Futures

Pour surmonter ces défis, certains auteurs [93] soulignent que l'apprentissage machine offre des opportunités significatives pour affiner la segmentation, notamment grâce à l'intégration de données en temps réel et à l'utilisation d'algorithmes avancés tels que les réseaux de neurones. Les perspectives futures incluent une personnalisation accrue des services et une réactivité améliorée face aux comportements des consommateurs.

D'autres études [94] proposent une approche novatrice pour la segmentation des clients en utilisant des techniques d'intelligence artificielle. Elles discutent également de l'importance de l'analyse des données et de l'intégration de méthodes avancées pour améliorer la précision de la segmentation et s'adapter aux comportements changeants des consommateurs, ce qui est crucial dans un environnement de marché dynamique.

Par ailleurs, certains chercheurs [95] explorent les applications de l'apprentissage machine pour la segmentation dynamique des clients, en particulier dans le secteur mobile. Ils abordent les défis et les opportunités associés à l'utilisation de données en temps réel pour améliorer les stratégies de segmentation.

En résumé, bien que le machine learning ait enrichi la segmentation des clients, des défis importants subsistent. Les recherches futures devront aborder la simplification des outils pour les PME, l'amélioration de l'adaptabilité des modèles, et la transition vers des techniques de segmentation dynamique.

4.2. Positionnement

La recherche actuelle en segmentation des clients par machine learning présente des lacunes, notamment le manque d'outils intégrés et conviviaux pour les petites et moyennes entreprises, une insuffisance d'études sur l'adaptabilité des modèles à divers contextes commerciaux, et une focalisation sur des segmentations statiques plutôt que dynamiques. Pour remédier à ces problèmes, notre étude se concentrera sur le développement d'un programme utilisant l'algorithme K-means comme solution pour la segmentation dynamique des clients. Ce programme sera conçu pour être performant et accessible à toutes les entreprises, tout en explorant des algorithmes facilement adaptables à différents secteurs et types de données. De plus, il proposera des solutions pour une segmentation dynamique en temps réel, permettant aux entreprises de réagir rapidement aux changements de comportement des clients et d'ajuster leurs stratégies en conséquence.



Chapitre 5 : Etude de cas



6.1. Cahier des charges

6.1.1. Expression des besoins

Nous souhaitons mettre en place :

- **Une segmentation démographique** : c'est le fait de répartir la population en groupes uniformes en fonction de caractéristiques telles que l'âge, le sexe, le revenu, le niveau d'éducation, la profession, etc., afin de mieux comprendre les besoins et les comportements des différents segments. Nous allons segmenter les clients suivant l'Age, le sexe et/ou l'Etat. Ainsi nous allons sortir les graphiques, les clusters et les prédictions suivantes :
 - Segmentation des clients suivant l'âge ;
 - Segmentation des clients suivant le genre ;
 - Segmentation des clients par Etat ;
 - Segmentation des clients par âge et genre ;
 - Segmentation des clients par âge et Etat ;
 - Segmentation des clients par genre et Etat ;
 - Segmentation des clients par âge, genre et Etat ;
 - Prédiction du genre du prochain client ;
 - Prédiction de l'âge du prochain client ;
 - Prédiction de l'Etat du prochain client ;
 - Prédiction du groupe (âge, Etat et genre) du prochain client.
- **Une segmentation transactionnelle** : c'est le fait de diviser les clients en groupes selon leurs comportements et leurs habitudes d'achat pour mieux leur proposer des produits, des offres ou des services. Nous allons segmenter les clients suivant la marque, la commande en ligne (si c'est commande en ligne ou pas) et/ou le prix de vente. Ainsi nous allons sortir les graphiques, les clusters et les prédictions suivantes :
 - Segmentation des transactions par Marque ;
 - Segmentation des transactions par commande en ligne ;
 - Segmentation des transactions par Prix de vente ;
 - Segmentation des transactions par marque et commande en ligne ;
 - Segmentation des transactions par commande en ligne et prix de vente ;
 - Segmentation des transactions par marque, commande en ligne prix de vente ;
 - Prédiction de la marque, de la commande en ligne et du prix de vente de la prochaine transaction.



6.1.2. Présentation du corpus d'étude

Notre étude de cas portera sur un jeu de données clients très populaire et très complet proposer par Sprocket Central Pty Ltd⁴¹, une organisation de taille moyenne spécialisée dans le domaine des vélos et des accessoires de cyclisme. Le jeu de données est disponible en téléchargement libre sur le site de KAGGLE⁴². L'organisation dispose d'un vaste ensemble de données relatives à ses clients qui peuvent permettre d'analyser efficacement ces données afin d'optimiser sa stratégie marketing. Le Dataset est dans un seul fichier Excel nommé KPMG_VI_New_raw_data_update_final.xlsx composé de cinq feuilles (voir le tableau et les figures suivants).

Tableau 2:Présentation des données Dataset

Données	Descriptions	Nombre de ligne	Nombre de colonne
Customer_transaction_dataset	Ce jeu de données contient des transactions client anonymisées d'un site e-commerce. Il peut inclure des données sur les achats, les comportements d'achat, les profils client, etc.		
Transactions	Contient les données transactions des clients	20 001	13
CustomerDemographic	Contient les données démographiques des clients	4 000	12
NewCustomerList	Contient la liste des nouveaux clients	1 000	23
Custom Adress	Contient l'Adresse des clients	4000	6

6.1.3. Choix des outils de travail

Le tableau suivant montre une description des outils de travail choisis.

⁴¹ <https://medium.com/@adetunjimammed26/sprocket-central-pty-ltd-a-bike-sales-company-analysis-fd868e0634fd>

⁴² <https://www.kaggle.com/datasets/archit9406/customer-transaction-dataset>



Environnement	Anaconda
Editeur de code	Jupyter Notebook
Langage	Python
SE	Windows
Bibliothèques de traitement de données	Pandas : Pour la manipulation et l'analyse des données. CSV : Pour la lecture et l'écriture de fichiers CSV. NumPy : Elle est souvent utilisée pour les opérations mathématiques sur les données.
Bibliothèques de segmentation des clients	Scikit-learn (sklearn) : elle est utilisée pour les algorithmes de segmentation comme KMeans et DBSCAN. sklearn. cluster. KMeans, sklearn. preprocessing. StandardScaler
Bibliothèques de machine learning	sklearn (scikit-learn) : c'est la bibliothèque principale pour les algorithmes de machine learning ; sklearn.preprocessing.LabelEncoder, sklearn.linear_model.LinearRegression, sklearn.neural_network.MLPClassifier, sklearn.ensemble.GradientBoostingClassifier, sklearn.ensemble.RandomForestRegressor
Bibliothèques de visualisation	Plotly

Tableau 3: Outils choisis

6.1.4. Installation des outils et bibliothèques

Des guides d'installations des outils choisis sont disponibles et recensés dans les tableaux suivants.

Outils	Lien guide d'installation
Pandas	pandas - Bibliothèque d'analyse de données Python (pydata.org)
Numpy	NumPy - Installation de NumPy
Scikit-Learn	Installation de scikit-learn — documentation de scikit-learn 1.5.0
Plotly	Premiers pas avec plotly en Python

Tableau 4: Guide d'installation des outils et API



6.2. Implémentation

6.2.1. Nettoyage des données

Nous avons choisi :

- De ne maintenir que les feuilles sur les transactions et la démographique. En effet, avec la feuille transactions qui contient les données transactionnelles tels que l'identifiant de transaction, l'identifiant de produit et l'identifiant de client, la date de la transaction, la commande en ligne, le statut de la commande, la marque, la ligne de produit, la classe de produit, la taille du produit, le prix de vente, le coût standard et la date de premières ventes ;
- Dans la feuille liste des nouveaux clients on a le prénom, le nom de famille, le genre, les achats liés aux vélos des trois dernières années, la date de naissance, le titre du poste, la catégorie d'industrie, le segment de richesse, l'indicateur de décès, la possession d'une voiture, l'ancienneté, l'adresse, le code postal, l'état, le pays, l'évaluation de la valeur du bien, le rang et la valeur ;
- Dans les données démographiques des clients on a l'identifiant client, le nom, le genre, les achats liés aux vélos des trois dernières années, la date de naissance, l'âge, le titre du poste, la catégorie d'industrie, le segment de richesse, l'indicateur de décès, la possession d'une voiture et l'ancienneté ;
- Et enfin dans l'adresse des clients on a l'identifiant client, l'adresse, le code postal, l'état, le pays et l'évaluation de la valeur du bien ;

6.2.2. Prétraitement des données

- Après cela, on va transformer les fichiers en dataset CSV pour une meilleure manipulation et réduire la dimension du fichier en fonction de nos besoins c'est-à-dire qu'on va choisir des critères dont on a seulement besoin comme dans :
 - o **Transactions** : on a choisi les données dont on aura seulement besoin pour la segmentation et la prédiction à savoir la marque, les commandes en ligne ((vrai=fait en ligne) ou (faux=présentielle)), et le Prix de vente ;
 - o **Infos _clients** : on a choisi les données dont on aura seulement besoin pour la segmentation et la prédiction à savoir le genre, l'âge, l'Etat.

6.2.3. Fonctionnement du programme

Le processus commence par charger les données du dataset, sélectionner les variables nécessaires, appliquer l'algorithme K-means pour segmenter les variables et se termine par la visualisation des résultats avec un graphique.



6.3. Interprétation des résultats

Après implémentation nous avons comme résultats :

- Segmentation des clients suivant le tranche d'âge ;

Segmentation des Clients par tranche d'âge avec K-means

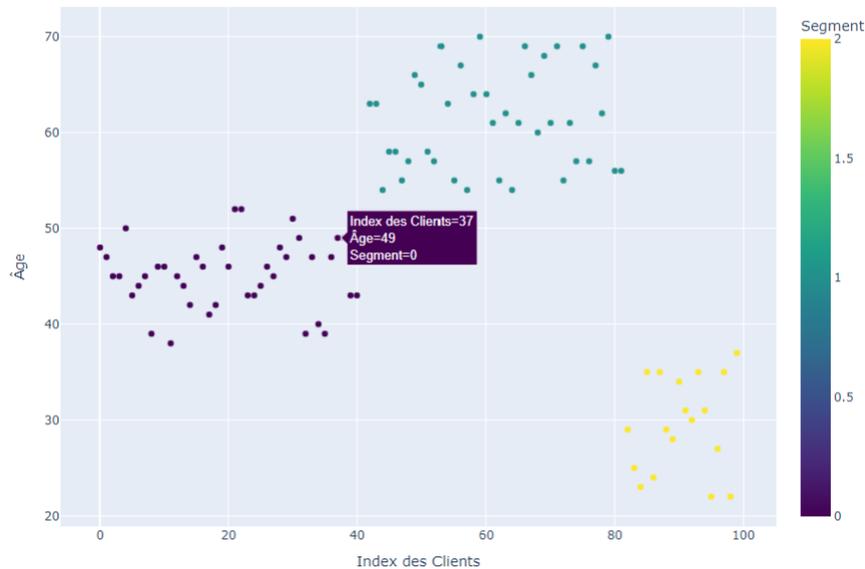


Figure 8: Graphique de la segmentation des clients par âge.

Ce nuage de points montre la segmentation des clients en trois groupes distincts en fonction de leur Âge. Chaque point représente un client, et les couleurs indiquent le segment auquel le client appartient. Ce graphique permet d'observer comment les clients se répartissent selon leur Âge dans les différents segments identifiés par K-means.

- Segmentation des clients suivant le Genre ;

Segmentation des Clients par Genre avec K-means

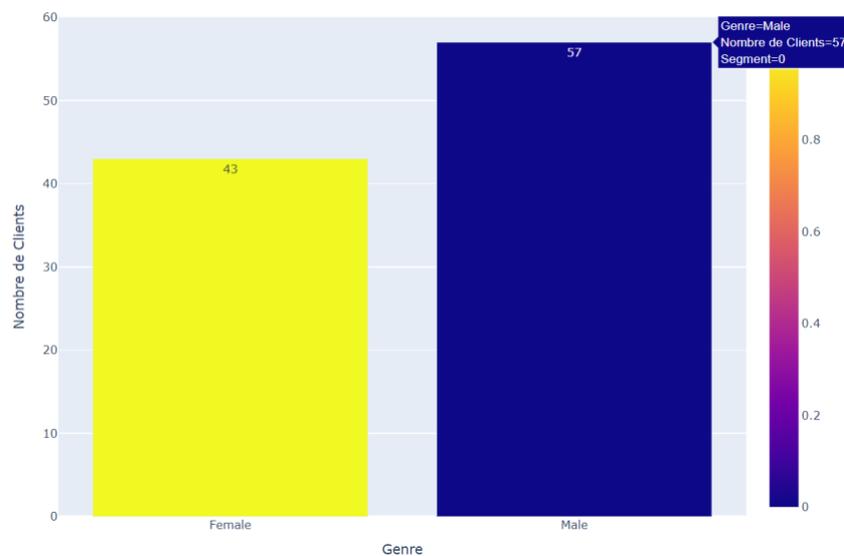


Figure 9: Graphique de la segmentation des clients par Genre.



Ce graphique en barres 2D présente la segmentation des clients en deux groupes distincts (Segments 0 et 1) en fonction de leur Genre, en utilisant l'algorithme K-means. Ce graphique montre comment les clients sont répartis entre les deux segments pour chaque Genre et permet de comparer le nombre de clients entre ces segments.

- Segmentation des clients par État ;

Segmentation des Clients par État avec K-means

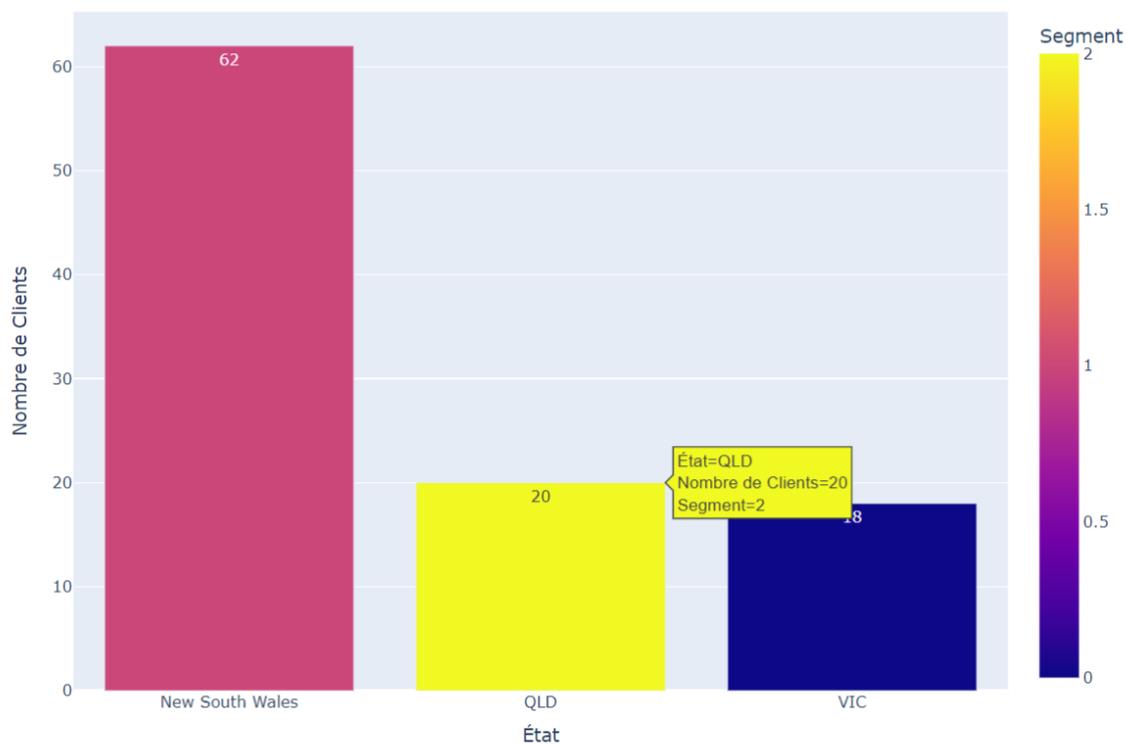


Figure 10: Graphique de la segmentation des clients par État.

Ce graphique en barres 2D présente la segmentation des clients en trois groupes distincts en fonction de l'État des clients, en utilisant l'algorithme K-means.

Ce graphique permet de visualiser la répartition des clients dans chaque État pour chaque segment créé par le modèle de clustering.

- Segmentation des clients par âge et genre ;

Le diagramme en barres suivant montre la segmentation des clients en deux groupes distincts en fonction de leur Âge et Genre. Chaque facette du graphique représente un Genre, et les barres montrent le nombre de clients dans chaque groupe d'Âge pour chaque Segment. Le graphique permet de comparer la répartition des clients en fonction de l'Âge entre les deux Genres et d'observer les caractéristiques de chaque Segment.



Segmentation des Clients par Âge et Genre avec K-means

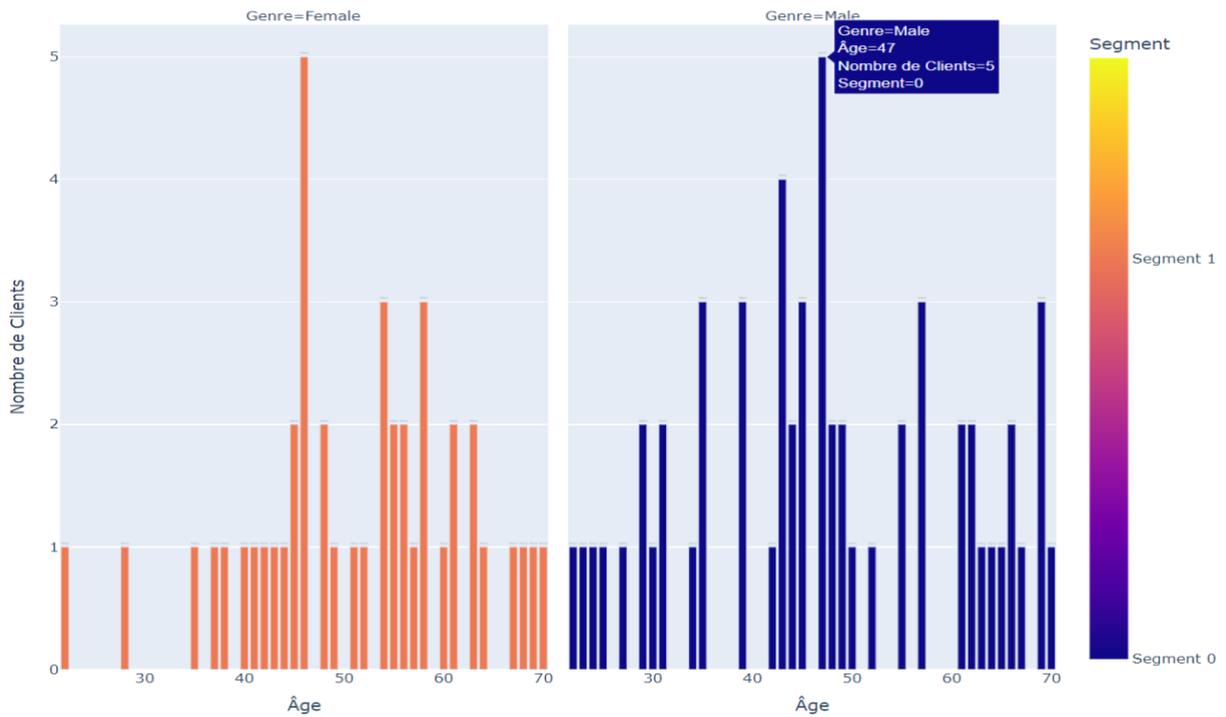


Figure 11: Graphique de la segmentation des clients par âge et genre.

- o Segmentation des clients par âge et Etat ;

Segmentation des Clients par Âge et État avec K-means

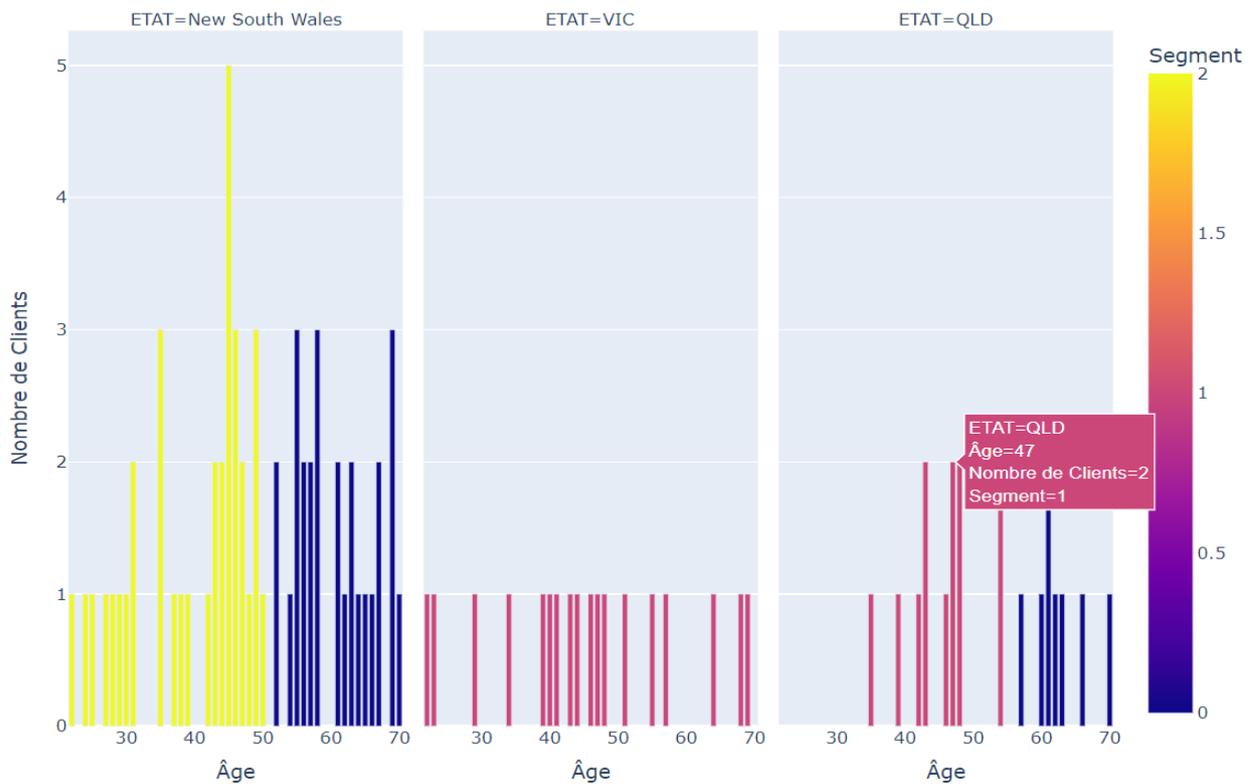


Figure 12: Graphique de la segmentation des clients par âge et Etat.

Ce diagramme en barres visualise la segmentation des clients en plusieurs groupes en fonction de leur Âge et de leur État. Chaque facette du graphique représente un État spécifique, et les barres montrent le nombre de clients dans chaque groupe d'Âge pour chaque Segment. Ce graphique permet de comparer la répartition des clients selon les âges entre différents États et d'observer les caractéristiques de chaque Segment.

- Segmentation des clients par genre et Etat ;

Segmentation des Clients par Genre et État avec K-means

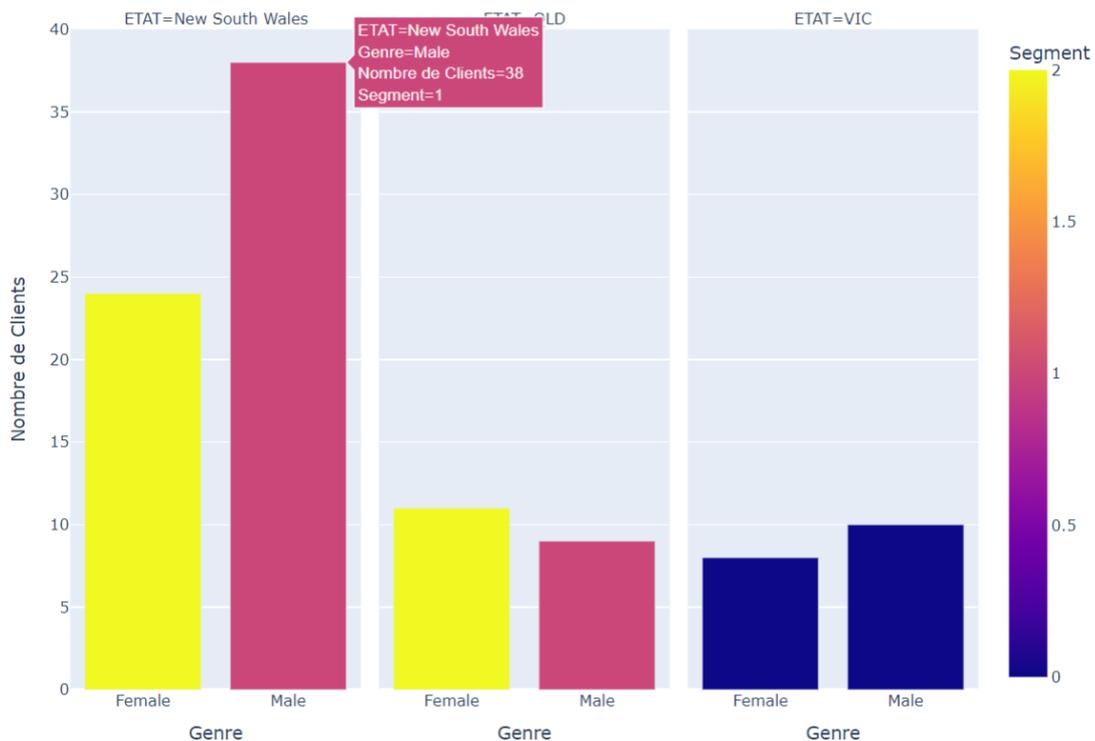


Figure 13: Graphique de la segmentation des clients par genre et Etat.

Ce diagramme en barres montre la segmentation des clients par genre et état en utilisant l'algorithme K-means. Chaque facette du graphique correspond à un état, avec des barres montrant le nombre de clients par genre et segment. Ce graphique facilite la comparaison des clients entre différents genres et états ainsi que l'analyse des caractéristiques des segments.

- Segmentation des clients par âge, genre et Etat ;

Le diagramme en barres suivant offre une vue d'ensemble claire de la répartition des clients dans différents segments en fonction de leur âge, genre et état. Chaque facette du graphique permet de visualiser le nombre de clients par tranche d'âge, avec des barres colorées pour chaque segment. Ce visuel aide à comprendre les différences et similitudes entre les clients en fonction de ces variables clés.

Segmentation des Clients par Âge, Genre et État avec K-means

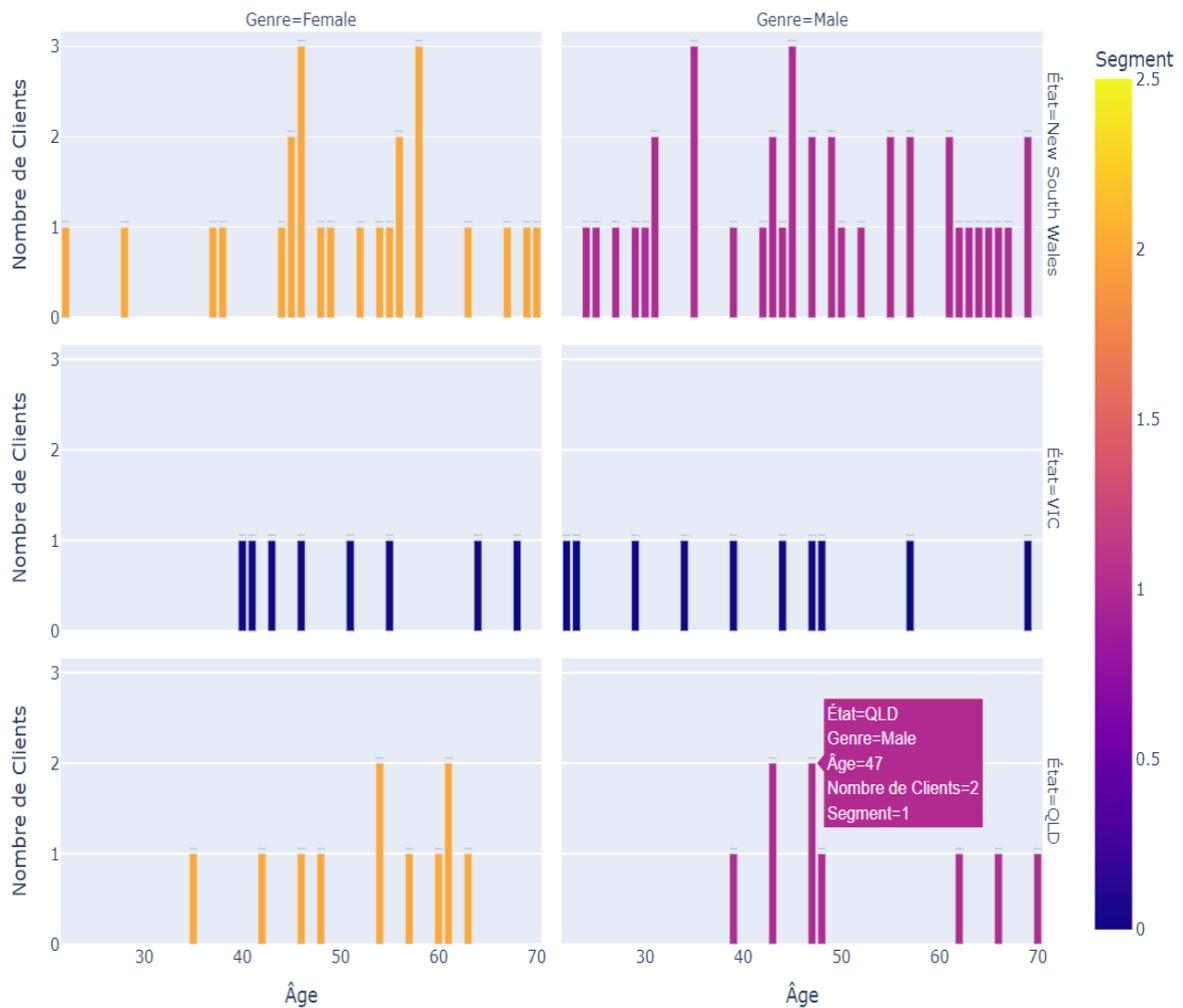


Figure 14: Graphique de la segmentation des clients par âge, genre et Etat.

- Prédiction du Genre du prochain client ;

Le graphique 3D suivant illustre la répartition des genres des clients en fonction de leur âge, prédite à l'aide d'un modèle de réseau de neurones multicouches (MLPClassifier), un algorithme d'apprentissage supervisé. Le modèle a été formé pour prédire le genre des clients en fonction de leur âge à partir de données historiques. Dans le graphique, les points colorés représentent les âges des clients, chaque couleur indiquant un genre différent (bleu pour Male, vert pour Female). Le point noir en forme de croix marque la prédiction du genre pour un client spécifique selon son âge. Ce graphique permet de visualiser les prédictions du modèle et d'évaluer sa performance par rapport aux données réelles.



Figure 15: Graphique de la prédiction du Genre du prochain client.

Le graphique 3D illustre la répartition des genres des clients en fonction de leur âge, prédite à l'aide d'un modèle de réseau de neurones multicouches (MLPClassifier), un algorithme d'apprentissage supervisé. Le modèle a été formé pour prédire le genre des clients en fonction de leur âge à partir de données historiques. Dans le graphique, chaque Etat est représentée d'une couleur. Le point noir en forme de croix marque la prédiction du genre pour un client spécifique selon son âge.

- Prédiction de l'âge du prochain client ;

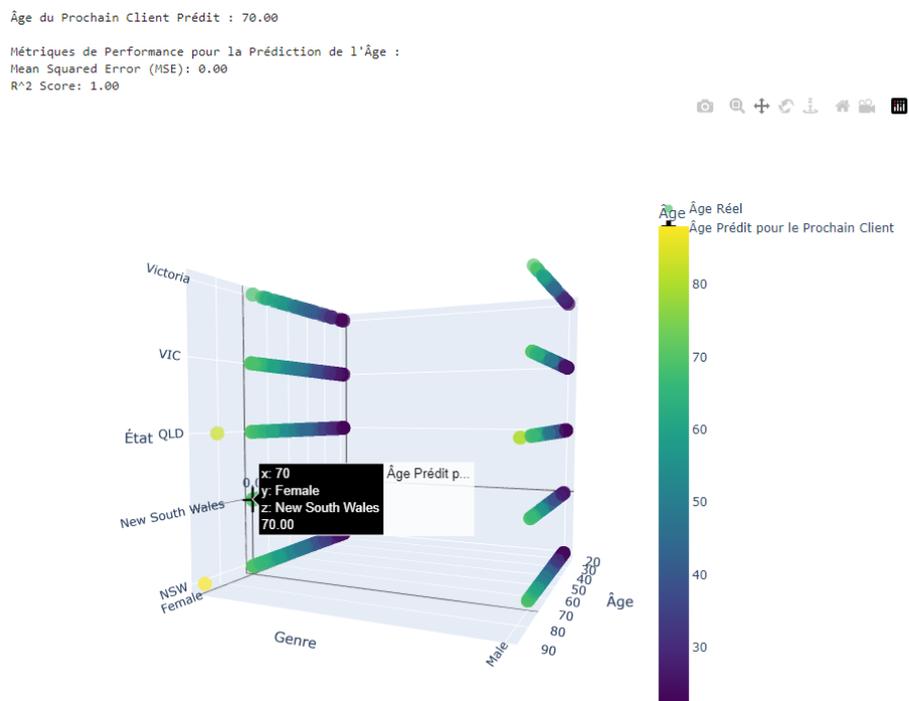


Figure 16: Graphique de la prédiction de l'âge du prochain client.



Le graphique 3D montre la prédiction de l'âge des clients en fonction de leur genre et de leur état, en utilisant un modèle de régression linéaire. Les points colorés représentent les âges réels des clients, avec une échelle allant du violet (jeune) au jaune (âgé). L'axe des y indique les genres et l'axe des z les états. Un point noir en croix montre l'âge prédit pour un client spécifique, basé sur ses caractéristiques. Ce graphique permet de visualiser la précision du modèle et de comparer les âges réels à l'âge prédit.

- Prédiction de l'Etat du prochain client ;

Le graphique 3D suivant illustre les prédictions d'un modèle de MLPClassifier pour déterminer l'état des clients en fonction de leur âge et genre. Les points colorés représentent les états réels des clients, tandis que les symboles montrent les états. Un point rouge en croix indique la prédiction du genre du prochain client. Ce graphique permet de visualiser et de comparer facilement la prédiction du modèle avec les états réels, tout en fournissant une évaluation de la performance du modèle à travers des métriques comme la précision et le rappel.

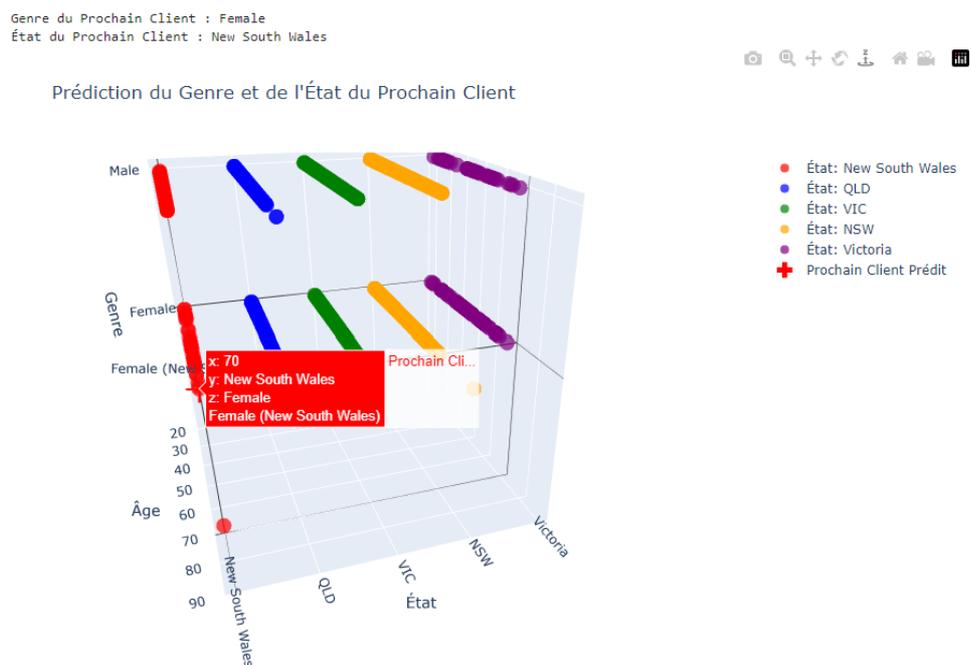


Figure 17: Graphique de la prédiction de l'Etat du prochain client.

- Prédiction de l'âge, Etat et le genre du prochain client ;

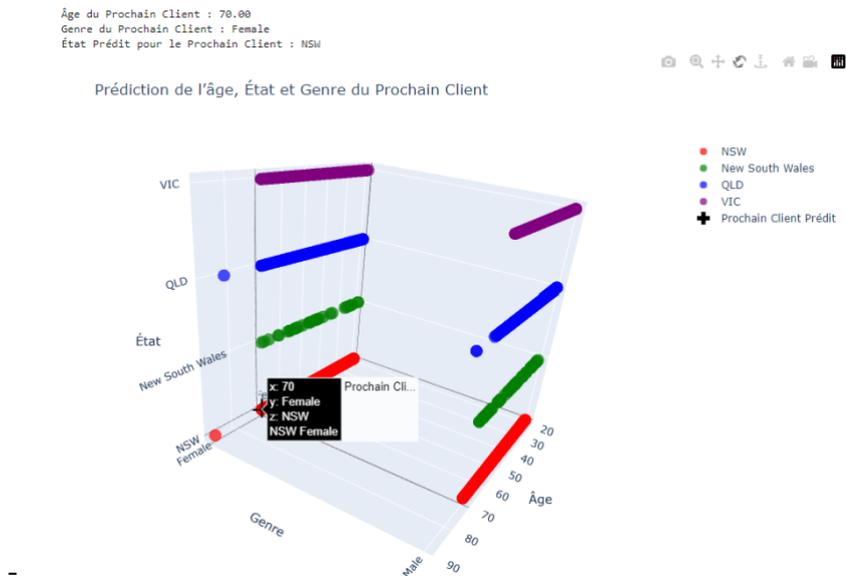


Figure 18: Graphique de la prédiction de l'âge, Etat et le genre du prochain client.

Le graphique 3D affiche les prédictions d'âge, de genre et d'état d'un client en utilisant divers modèles d'apprentissage supervisé : régression linéaire pour l'âge, MLPClassifier pour le genre et l'Etat. Les données réelles sont représentées par des points colorés, tandis qu'un point noir en croix illustre les prédictions pour un client spécifique.

- **Une segmentation transactionnelle** : c'est le fait de diviser les clients en groupes selon leurs comportements et leurs habitudes d'achat pour mieux leur proposer des produits, des offres ou des services. Nous allons segmenter les clients suivant l'Age, le sexe et/ou le pays. Ainsi nous allons sortir les graphiques, les clusters et les prédictions suivantes :
 - o Segmentation des transactions par Marque ;



Segmentation des Transactions par Marque avec K-means

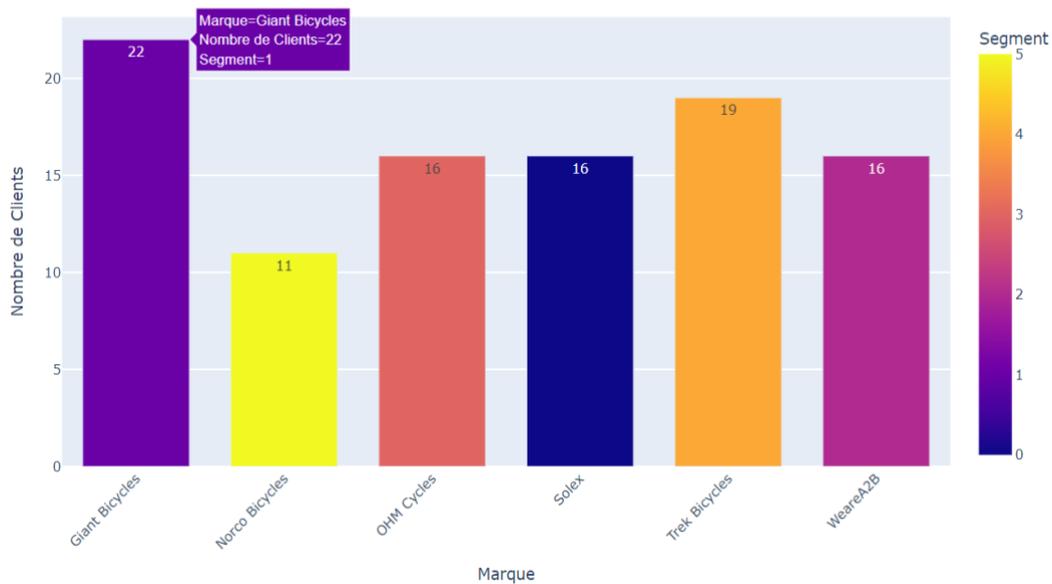


Figure 19: Graphique de la segmentation des transactions par Marque.

Ce graphique en barres 2D démontre la répartition des clients par marque et par segment K-means. Il aide à comprendre comment les différentes marques sont regroupées en segments et combien de clients se trouvent dans chaque segment pour chaque marque.

- Segmentation des transactions par commande en ligne ;

Ce graphique suivant permet de visualiser comment les commandes en ligne et hors ligne sont réparties parmi les différents segments de clients, identifiés par l'algorithme de clustering K-means. Il montre comment les clients se regroupent en fonction du prix de vente et du fait qu'ils ont passé une commande en ligne ou non.



Segmentation des Transactions par Commande en Ligne avec K-means

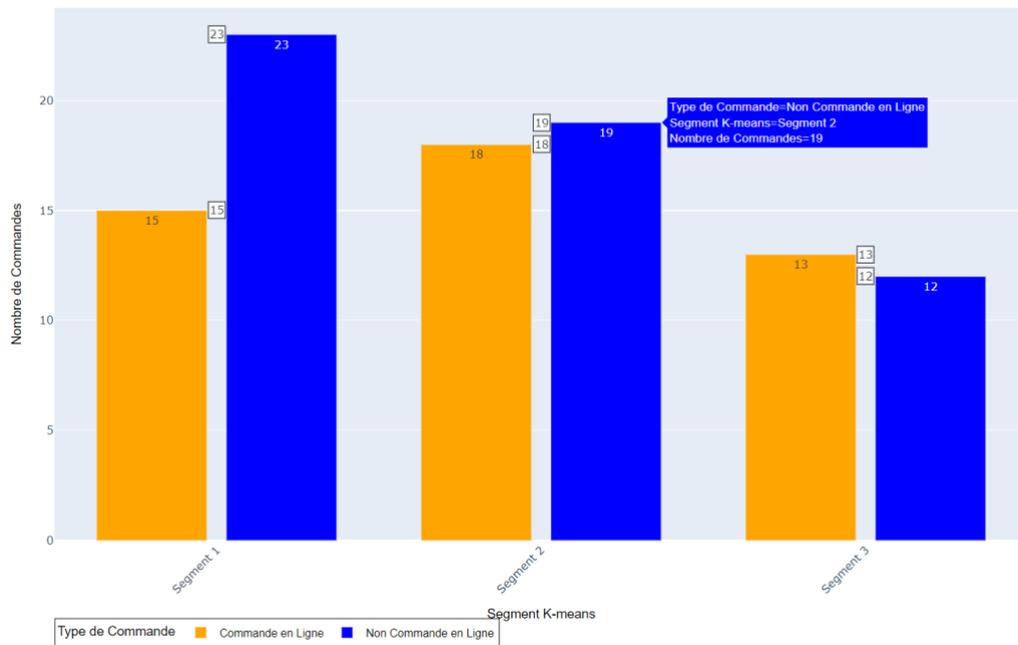


Figure 20: Graphique de la segmentation des transactions par commande en ligne.

- Segmentation des transactions par Prix de vente et commande en ligne ;

Segmentation des Transactions par Prix de Vente et Type de Commande avec K-means

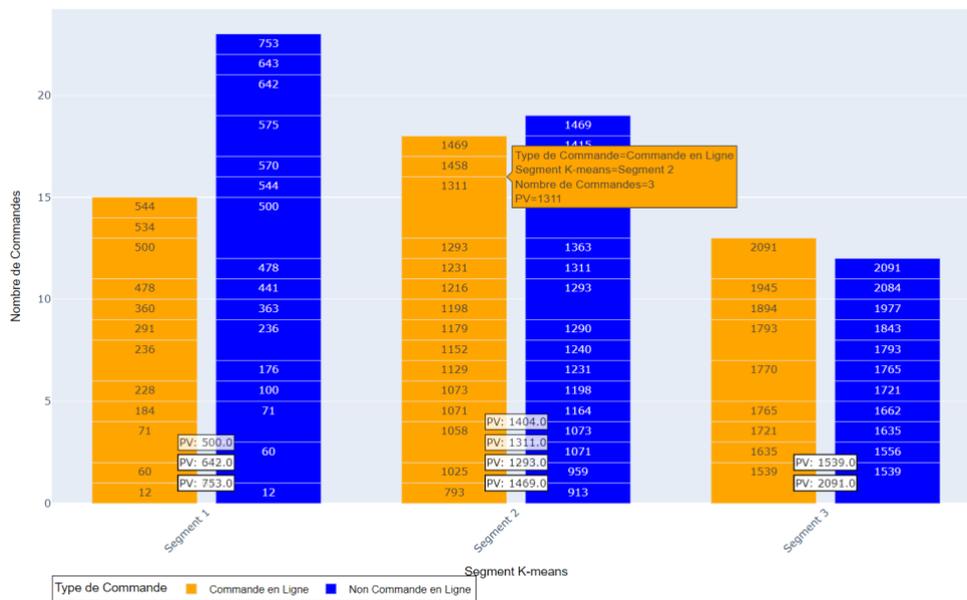


Figure 21: Graphique de la segmentation des transactions par Prix de vente et commande en ligne.

Ce graphique permet de visualiser la distribution des commandes en ligne et hors ligne à travers différents segments de clients, en prenant en compte le prix de vente des produits.

- Segmentation des transactions par marque et commande en ligne ;

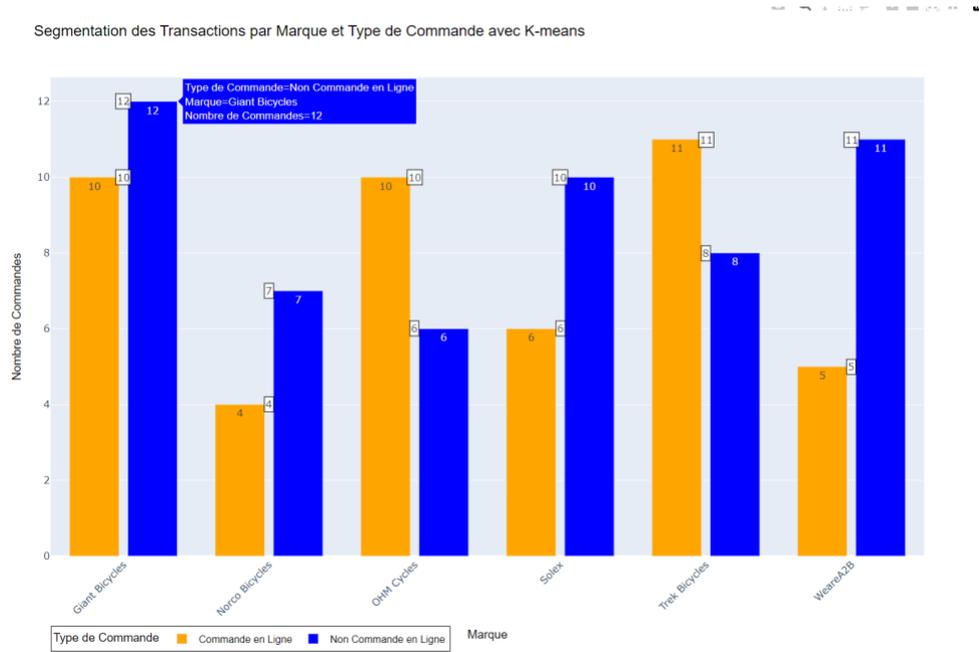


Figure 22: Graphique de la segmentation des transactions par marque et commande en ligne.

Ce graphique nous permet de visualiser le nombre de commandes en ligne et hors ligne pour différentes marques, en montrant comment ces commandes sont réparties entre ces deux catégories pour chaque marque.

Segmentation des Transactions par Marque et Prix de Vente avec K-means

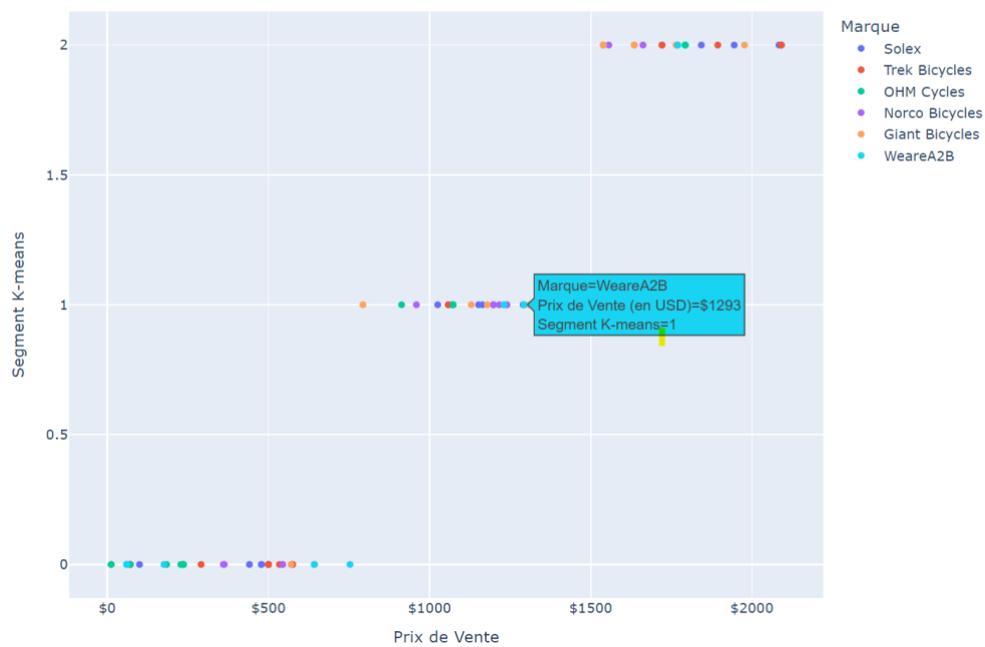


Figure 23: Graphique de la segmentation des transactions par marque et prix de vente.



Ce graphique 3D permet de visualiser la distribution des transactions selon les marques et les prix de vente, et comment ces transactions sont regroupées en segments distincts par l'algorithme de K-means.

- Segmentation des transactions par commande en ligne et prix de vente ;

Segmentation des Transactions par Commande en Ligne et Prix de Vente avec K-means

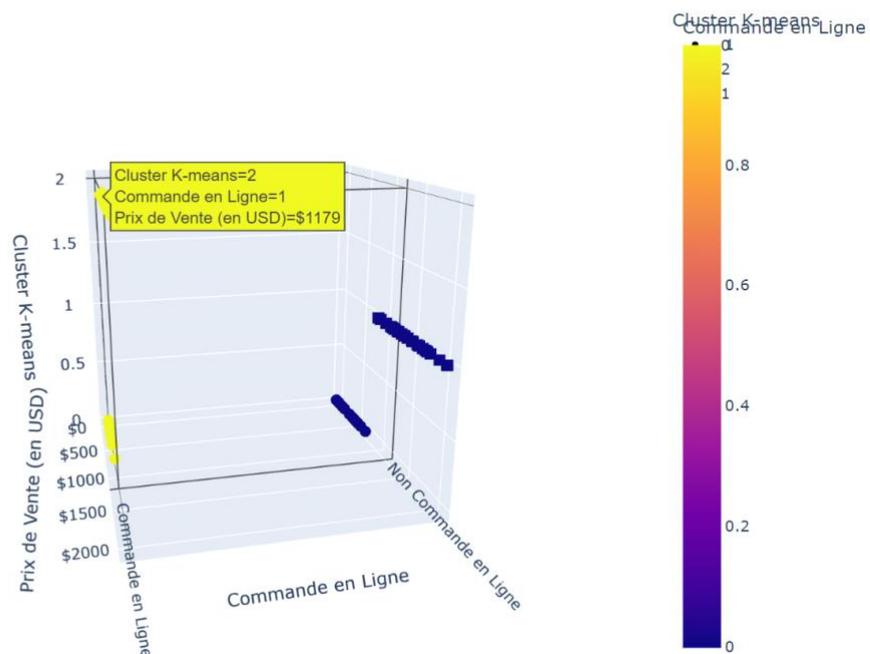


Figure 24: segmentation des transactions par commande en ligne et prix de vente

Ce graphique montre comment les transactions sont réparties selon qu'elles sont faites en ligne ou non, ainsi que leurs prix de vente. Il utilise des couleurs différentes pour distinguer les commandes en ligne (en jaune) et hors ligne (en bleu). En utilisant l'algorithme K-means, il identifie des groupes de transactions similaires, ce qui aide à comprendre les tendances des prix et des différents types de transactions.

- Segmentation des transactions par marque, commande en ligne et prix de vente ;

Ce graphique 3D montre la segmentation des transactions en fonction des marques, des types de commandes (en ligne ou hors ligne) et des prix de vente. Les points colorés représentent les transactions, avec des couleurs différentes pour chaque marque



Segmentation des Transactions par Commande en Ligne, Marque et Prix de Vente avec K

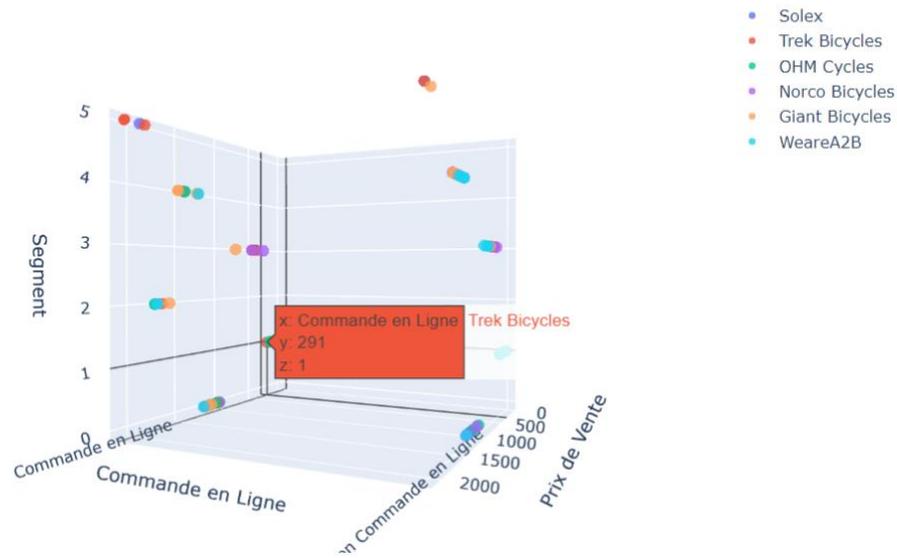


Figure 25: Graphique de la segmentation des transactions par marque, commande en ligne et prix de vente.

- Prédiction de la marque, commande en ligne et prix de vente de la prochaine transaction

Le graphique 3D suivant montre les marques actuelles et la transaction prédite, avec les marques sur l'axe des x, les prix de vente sur l'axe des y, et les commandes en lignes sur l'axe des z, chaque marque ayant une couleur distincte. La prédiction est marquée par un point violet et rouge, indiquant la marque, le statut de commande en ligne et le prix de vente estimé. Les modèles utilisés incluent le Gradient Boosting Classifier pour prédire la marque et la commande en ligne, et le Random Forest Regressor pour estimer le prix de vente,

Prediction du Marque, commande en ligne, et Prix de Vente du prochain client



Figure 26: Graphique de la prédiction de la marque, de la commande en ligne et du prix de vente de la prochaine transaction.



Conclusion et Perspectives

Dans ce mémoire, nous avons étudié l'utilisation du machine learning dans la segmentation des clients. En effet, la segmentation des clients par machine learning représente une approche stratégique essentielle dans le domaine du marketing, offrant une compréhension plus profonde et plus précise des comportements d'achat des clients. Grâce à une analyse avancée des données, cette pratique permet aux entreprises de personnaliser leurs offres et leurs stratégies de manière plus efficace, renforçant ainsi leur compétitivité sur le marché.

Les retombées de cette étude montrent que l'application du machine learning à la segmentation des clients permet non seulement de révéler des segments cachés et de faire des prédictions précises sur les clients et les transactions, mais aussi d'améliorer la personnalisation des offres et des stratégies marketing. Cela renforce la compétitivité des entreprises et la fidélité des clients.

Pour tirer pleinement parti des avantages de la segmentation des clients par machine learning, il est nécessaire que les entreprises continuent à investir dans la recherche et le développement de nouvelles techniques et technologies. En particulier, l'intégration de l'apprentissage automatique dans les systèmes de gestion de la relation client (CRM) permettrait une segmentation plus dynamique et en temps réel, offrant des opportunités d'interaction personnalisée avec les clients. De plus, en explorant des méthodes pour surmonter les défis liés à la confidentialité des données et à la gestion des biais algorithmiques, les entreprises peuvent garantir une utilisation éthique et transparente de la segmentation des clients par machine learning, renforçant ainsi la confiance des consommateurs et leur fidélité à long terme.

Nous avons mis en pratique le machine learning dans la segmentation en utilisant un jeu de données populaire contenant des données commerciales issues d'une boutique en ligne. Grâce à ces données, nous avons pu mettre en exergue des segments cachés et faire des prédictions sur les clients et les transactions.

Ce mémoire nous a également permis de nous familiariser avec le machine learning et son application dans un domaine en pleine expansion. Pour les travaux futurs, plusieurs perspectives s'ouvrent :

- Envisager l'application d'algorithmes de deep learning pour la segmentation des clients ;
- Explorer de nouvelles applications de l'IA générative pour la segmentation des clients ;
- Réaliser une étude comparative des approches supervisées, non supervisées et de deep learning pour évaluer leurs performances respectives dans la segmentation des clients ;
- Proposer un prototype d'outil pour la segmentation automatique des clients ;
- Proposer un système de recommandation.



Références

Articles scientifiques & livres

1. Kotler, P., Kartajaya, H., Setiawan, I., & Vandercammen, M. (2022). Chapitre 8. Le marketing piloté par les données. *Marketing*, 155-168
2. Bernelin, M. (2019). Intelligence artificielle en santé: la ruée vers les données personnelles. *Cités*, (4), 75-89.
3. Joliveau, T. (2011). Le géoweb, un nouveau défi pour les bases de données géographiques. *Espace géographique*, 40(2), 154-163.
4. Grolleau, E., Hugues, J., Yassine, O., & Henri, B. (2018). Introduction aux systèmes embarqués temps réel: Conception et mise en oeuvre. Dunod.
5. Madakam, S., Ramaswamy, R. et Tripathi, S. (2015). Internet des objets (IoT) : une revue de la littérature. *Journal de l'informatique et des communications*, 3(5), 164-173.
6. Toutou, C. (2021). Ce que la segmentation des publics apporte à l'amélioration des services en bibliothèque.
7. Mahesh, B. (2020). Algorithmes d'apprentissage automatique - une revue. *Revue internationale de la science et de la recherche (IJSR)*. [Internet], 9(1), 381-386.
8. Candillier, L. (2006). Contextualisation, visualisation et évaluation en apprentissage non supervisé (Doctoral dissertation, Université Charles de Gaulle-Lille III).
9. SENIHJI, K. (2023). L'INTELLIGENCE ARTIFICIELLE : ETATS DES LIEUX ET APPLICATIONS DANS LE DOMAINE DU MARKETING. *Revue d'Etudes en Management et Finance d'Organisation*, 8(2).
10. SOUADDA, L. I., BERGHOUT, Y. M., MENARI, M., & BENILLES, B. La gestion du risque des crédits d'exploitation des PME par le Machine Learning Étude de cas: La Banque CPA.
11. Crié, D., & Benavent, C. (1998). La dynamique de clientèle: segmentation et valeur actualisée du client. In *Valeur, marché et organisation*.
12. Kotler, P., Kartajaya, H., Setiawan, I., & Vandercammen, M. (2022). Chapitre 9. Le marketing prédictif. *Marketing*, 169-183.
13. SIBONY, A. L., & CLIFFORD, D. (2020). La personnalisation illicite. Florence G'ssell; «Le big data et le droit», 249-271.
14. Deveaux, L. (2005). Le rôle stratégique des systèmes de tarification dynamique sur l'Internet. *Systèmes d'Information et Management*, 10(2), 63.



15. Omrani, N., Omrani, T., & Rallet, A. (2023). Une analyse des comportements de recommandation de produits à partir des avis postés sur Amazon. *Innovations*, I157-XXXIV.
16. LEONARD, T. (2024). La compréhension et la prédiction des préférences des clients dans le commerce en détail grâce à l'IA: Une revue de littérature. *International Journal of Economic Studies and Management (IJESM)*, 4(1), 95-107.
17. Hadjitchoneva, J. (2020). L'intelligence artificielle au service de la prise de décisions plus efficace. Pour une recherche économique efficace, 149.
18. Li, Y. et Wu, H. (2012). Une méthode de clustering basée sur l'algorithme K-means. *Physics Procedia*, 25, 1104-1109.
19. Berger, J. L. (2021). Analyse factorielle exploratoire et analyse en composantes principales: guide pratique.
20. BOUCHICHA, I., DIF Roumaissa, N. E. Y., & CHERAITIA, H. E. (2018). Analyse factorielle discriminante (Doctoral dissertation, Université jijel).
21. Tufféry, S. (2009). Etude de cas en statistique décisionnelle. Editions Technip.
22. Arefin, S., Parvez, R., Ahmed, T., Ahsan, M., Sumaiya, F., Jahin, F., & Hasan, M. (2024, May). Retail Industry Analytics: Unraveling Consumer Behavior through RFM Segmentation and Machine Learning. In 24th Annual IEEE International Conference on Electro Information Technology (eit2024). Consulté le 24/07/2024
23. Maričić, B. R., & Đorđević, A. Članci/Articles.
24. Jourani, R., Daoudi, K., André-Obrecht, R., & Aboutajdine, D. (2011, September). Apprentissage discriminant des GMM à grande marge pour la vérification automatique du locuteur. In GRETSI.
25. Agrawal, A., Kaur, P., & Singh, M. (2023, mars). Modèle de segmentation de la clientèle utilisant le clustering K-means sur le commerce électronique. Dans la Conférence internationale 2023 sur les systèmes informatiques et de communication de données durables (ICSCDS) (pp. 1-6). IEEE.
26. Kamthania, D., Pawa, A., & Madhavan, S. S. (2018). Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business. *Journal of computing and information technology*, 26(1), 57-68.
27. Hossain, AS (2017, décembre). Segmentation de la clientèle à l'aide d'algorithmes de clustering basés sur le centroïde et la densité. En 2017, 3e Conférence internationale sur les technologies de l'information et de la communication électriques (EICT) (pp. 1-6). IEEE.



28. Hicham, N., Karim, S., & Habbat, N. (2022, November). Novel Approach and Innovative Strategy for Mall Customer Segmentation Using Machine Learning Techniques. In International Conference on Advanced Technologies for Humanity (pp. 43-52). Cham: Springer Nature Switzerland.
29. Baray, J., Bécheur, A., & Soulabail, Y. (2022). Appréhender les thématiques et communautés de recherche en gestion à travers l'analyse sémantique des publications en sciences de gestion sur HAL. *La Revue des Sciences de Gestion*, (1), 13-24.\
30. Guerradi, N., & Benkhelifa, H. (2019). Reconnaissance des émotions faciales par apprentissage profond (Doctoral dissertation, جامعة غرداية).
31. Philippot, E., Belaïd, Y., & Belaïd, A. (2010, March). Classification de formulaires manuscrits en-ligne à l'aide de réseaux bayésiens. In Colloque International Francophone sur l'Ecrit et le Document-CIFED 2010 (pp. 95-110).
32. Djuriscic, V., Kascelan, L., Rogic, S., & Melovic, B. (2020). Optimisation de la gestion de la relation client (CRM) des banques par classification prédictive basée sur la méthode de la machine à vecteurs de support. *Applied Artificial Intelligence*, 34 (12), 941-955.
33. Chiedozié, OJ, Umar, MI, & Chukwunwike, AJ (2021). Segmentation de la clientèle à l'aide de la technique d'exploration de données : une étude de cas sur les industries de vente au détail au Nigéria.
34. Ganar, C., & Hosein, P. (2022, décembre). Segmentation de la clientèle pour améliorer les campagnes marketing dans le secteur bancaire. En 2022, 5e conférence asiatique sur l'apprentissage automatique et l'informatique (ACMLC) (pp. 48-52). IEEE.
35. Cuxac, P., & Lamirel, J. C. (2011). Clustering incrémental et méthodes de détection de nouveauté: application à l'analyse intelligente d'informations évoluant au cours du temps. *La recherche d'information en contexte: Outils et usages applicatifs*.
36. Moins, T. (2020). *Modèle Hybride Combinant Réseau De Neurones Convolutifs et Modèle Vase Sur Le Choix Pour La Recommandation De Sièges*. Ecole Polytechnique, Montreal (Canada).
37. Hartert, L. (2010). Reconnaissance des formes dans un environnement dynamique appliquée au diagnostic et au suivi des systèmes évolutifs (Doctoral dissertation, Université de Reims-Champagne Ardenne).
38. Chucya Lozano, W. K. (2020). Développement d'un modèle de prédiction de retours de vêtements en fonction des caractéristiques des clients et des produits (Doctoral dissertation, Polytechnique Montréal).



39. Hassan, M. M. T. M., & Tabasum, M. (2018). Customer profiling and segmentation in retail banks using data mining techniques. *International journal of advanced research in computer science*, 9(4), 24-29.
40. Pondel, M., & Korczak, J. (2018, September). Collective clustering of marketing data-recommendation system upsaily. In *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 801-810). IEEE.
41. La Cruz, A., Severeyn, E., Matute, R., & Estrada, J. (2021). Segmentation des utilisateurs basée sur les revenus de Google Analytics à l'aide de k-means. Dans *Technologies de l'information et de la communication : 9e Conférence de l'Équateur, TICEC 2021, Guayaquil, Équateur, 24-26 novembre 2021, Actes 9* (pp. 225-235). Springer International Publishing.
42. LANZUISE, V. (2017). Clusterisation des utilisateurs et personnalisation de l'expérience utilisateur dans l'analyse Web : mise en œuvre et surveillance des KPI de retargeting.
43. Hemadharshini, SM, Kanchana Devi, R., Rajakumari, S., et Adline Freeda, R. (2023, juin). Segmentation des clients du commerce électronique par apprentissage non supervisé. Dans *Conférence internationale sur le calcul souple et le traitement du signal* (pp. 285-296). Singapour : Springer Nature Singapore.
44. Bartholomeus, C., & Kervyn de Meerendré, N. Sauver la bio à l'aide du Marketing Automation Analyse et recommandations pour le cas de Färm: Le Marketing Automation comme outil de fidélisation.
45. Zhuang, QR, Yao, YW et Liu, O. (2018). Application de l'exploration de données dans le marketing des dépôts à terme. *iaeng. org* .
46. Julian, A., et Hariprasath, SR (2024, février). Optimisation de la segmentation des clients grâce à l'apprentissage automatique. *Conférence internationale IEEE 2024 sur les technologies de l'informatique, de l'énergie et de la communication (IC2PCT)* (vol. 5, pp. 413-416). IEEE.
47. Géron, A. (2022). *Apprentissage automatique pratique avec Scikit-Learn, Keras et TensorFlow* . " O'Reilly Media, Inc. " .
48. Nayak, R., Dinesh, S., et Thirunavukkarasu, S. (2019). Une nouvelle méthode d'amélioration du mineur rapide pour les applications d'exploration de données. *Revue internationale des technologies innovantes et de l'ingénierie d'exploration* , 8 (9), 457-460.
49. Iqbal, MA SEGMENTATION CLIENTÈLE PAR CLASSIFICATION À L'AIDE DE WEKA. « *INFORMATIQUE MODERNE ET SES MÉTHODES D'ENSEIGNEMENT (MITM2020)* , 50.



50. Ahmed, I., House, R., Deilma, N., & Luo, L. (2019, novembre). Modèle de reconnaissance visuelle personnalisé avec Watson Studio. Dans les actes de la 29e conférence internationale annuelle sur l'informatique et le génie logiciel (pp. 376-377).
51. Barga, R., Fontama, V., Tok, WH, et Cabrera-Cordon, L. (2015). Analyse prédictive avec l'apprentissage automatique Microsoft Azure (pp. 221-241). Berkely, CA : Apress.
52. Wang, X. (2016). Apprentissage profond dans la reconnaissance, la détection et la segmentation d'objets. *Foundations and Trends® in Signal Processing* , 8 (4), 217-382.
53. Perdana, A., et Wang, T. (2023). Exploiter les données et adopter une stratégie numérique chez Zalora. *Journal of Emerging Technologies in Accounting* , 20 (1), 111-123.
54. Kodali, R. K., Panda, A., & Boppana, L. (2023, October). Attendance System using Amazon Rekognition. In *TENCON 2023-2023 IEEE Region 10 Conference (TENCON)* (pp. 65-70). IEEE.
55. Fulzele, V., Kirad, P., Dubey, C., Kulkarni, Y., & Thatte, S. (2021, mai). Utilisation des capacités du cloud pour la détection et la reconnaissance faciales pendant la COVID-19 : analyse comparative. Dans les actes de la Conférence internationale sur l'intelligence des données intelligentes (ICSMDI 2021) .
56. Eddine, T. H. (2020). La détection d'objet avec OpenCV et deep learning.
57. HAFIANE, A., & HADJERCI, O. La détection de tête en temps réel dans les scènes localement dense à l'aide des dernières techniques de Deep Learning.
58. MESBAH, F. (2021). Détection d'objets par Deep Neural Network à l'aide du modèle YOLO en temps réel.
59. Minkesh, A., Worranita, K., et Taizo, M. (2019). Extraction humaine et transition de scène utilisant Mask R-CNN. Préimpression arXiv arXiv:1907.08884 .
60. Wang, Z., Wang, J., Yang, K., Wang, L., Su, F., & Chen, X. (2022). Segmentation sémantique d'images de télédétection haute résolution basée sur un mécanisme d'attention aux caractéristiques de classe fusionné avec Deeplabv3+. *Computers & Geosciences* , 158 , 104969.
61. Busquet, F. De KITT à Amazon Alexa : une introduction à la reconnaissance automatique de la parole et à la synthèse vocale avec des applications en marketing.
62. Boonstra, L. (2021). Guide définitif de l'IA conversationnelle avec dialogflow et Google Cloud (p. 408). Berkeley : Apress.
63. Hegdepatil, P., & Davuluri, K. (2021, janvier). Approche stratégique marketing innovante basée sur la veille stratégique utilisant la reconnaissance vocale automatique et la synthèse



- de texte. En 2021, 2e Conférence internationale sur l'informatique et la science des données (CDS) (pp. 595-602). IEEE.
64. Moniz, A., Gordon, M., Bergum, I., Chang, M., Grant, G., Moniz, A., ... & Grant, G. (2021). Présentation des services cognitifs. Début des services cognitifs Azure : prise de décision basée sur les données grâce à l'intelligence artificielle , 1-17.
 65. Tang, T., Huang, L., & Chen, Y. (2020, December). Evaluation of Chinese sentiment analysis APIs based on online reviews. In 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 923-927). IEEE.
 66. Terasaka, DT, Martins, LE, dos Santos, VA, Ventura, TM, de Oliveira, AG et de SG Pedroso, G. (juillet 2024). Segmentation audio pour créer des ensembles de données sur le dressage des oiseaux. Dans Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA) (pp. 199-202). SBC.
 67. Medvedev, MS, et Okuntsev, YV (2019, décembre). Utilisation de la bibliothèque d'apprentissage automatique Google tensorFlow pour la reconnaissance vocale. Dans Journal of Physics : Conference Series (Vol. 1399, No. 3, p. 033033). IOP Publishing.
 68. Mizera, P., & Pollak, P. (2018). Segmentation phonétique automatique et détection de prononciation avec diverses approches de modélisation acoustique. Dans Speech and Computer : 20e Conférence internationale, SPECOM 2018, Leipzig, Allemagne, 18-22 septembre 2018, Actes 20 (pp. 419-429). Springer International Publishing.
 69. White, TE et Rege, M. (2020). Analyse des sentiments sur la plateforme Google Cloud. Issues in Information Systems , 21 (2), 221-228.
 70. Dingus, R. et Black, H. G. (2021). Choisissez vos mots avec soin : Un exercice pour introduire l'intelligence artificielle dans la classe de marketing à l'aide de l'analyse des tons. Revue de l'éducation en marketing, 31(2), 64-69.
 71. Roetzer, P. et Kaput, M. (2022). Intelligence artificielle marketing : IA, marketing et avenir des affaires. Livres BenBella.
 72. Barga, R., Fontama, V., Tok, W. H. et Cabrera-Cordon, L. (2015). Analytique prédictive avec l'apprentissage automatique de Microsoft Azure (pp. 221-241). Berkely, Californie : Avril.
 73. Padmaja, S., et Fatima, S. S. (2013). Exploration de l'opinion et analyse des sentiments - une évaluation de la croyance des gens : une enquête. Revue internationale d'informatique ad hoc, Capteurs et informatique ubiquitaire, 4(1), 21.



74. Tymann, K., Lutz, M., Palsbröcker, P. et Gips, C. (2019, septembre). GerVADER - Une adaptation allemande de l'outil d'analyse des sentiments VADER pour les textes des médias sociaux. Dans LWDA (pp. 178-189).
75. Gujjar, J. P. et Kumar, H. P. (2021). Analyse des sentiments : Textblob pour la prise de décision. *Int. J. Sci. Res. Eng. Trends*, 7(2), 1097-1099.
76. Adl, A., & Yacoup, K. M. Détermination de la satisfaction des clients des télécommunications à partir de leurs traits de personnalité à l'aide de la compréhension du langage naturel et de la SVM. Dans la quatrième conférence étudiante du secteur de l'éducation et des affaires étudiantes intitulée Innovation et créativité pour les étudiants de licence dans les universités égyptiennes, arabes et africaines.
77. Nirwana, A. (2023). Mise en œuvre de l'intelligence artificielle dans le développement du marketing numérique : une revue thématique et une exploration pratique. *Jurnal Manajemen Bisnis, Akuntansi Dan Keuangan*, 2(1), 85-112.
78. Tang, N. (2023, June). Analysis of Social Media Marketing Strategy: The Case Study of China and the US. In *Proceedings of the 6th International Conference on Economic Management and Green Development* (pp. 205-212). Singapore: Springer Nature Singapore.
79. MIMOUNE, Zakarya, OUAHAB, Abdelwhab, et al. Développement d'une Architecture Basée sur l'Apprentissage Profond (Deep Learning) pour la Détection d'Intrusion dans les Réseaux. 2019. Thèse de doctorat. Université Ahmed Draïa-Adrar.
80. Martín Gómez, S., et Muñoz de Luna, A. B. (2024). Le rôle des réseaux sociaux dans la communication dans le milieu de la recherche scientifique. *JOTSE : Journal de l'enseignement de la technologie et des sciences*, 14(2), 291-305.
81. Slavova, M. (2020). Identifying business needs. In *Digital and Social Media Marketing* (pp. 39-65). Routledge.
82. Tisserand-Barthole, C. (2015). MENTION, un outil de veille avec du potentiel. *Netsources*, (119), 9.
83. Munier, M., Lalanne, V., Ardoy, P. Y., & Ricarde, M. (2014, May). Métadonnées et Aspects Juridiques: Vie Privée vs Sécurité de l'Information. In 9ème Conférence sur la Sécurité des Architectures Réseaux et des Systèmes d'Information (SARSSI'2014) (pp. 65-76).
84. LeCun, Y. (2016). L'apprentissage profond, une révolution en intelligence artificielle. *La lettre du Collège de France*, (41), 13.



85. Wedel, M. (2000). *Segmentation du marché : Fondements conceptuels et méthodologiques*. Kluwer Academic Publisher.
86. Alves Gomes, M. et Meisen, T. (2023). Un examen des méthodes de segmentation de la clientèle pour le ciblage personnalisé des clients dans les cas d'utilisation du commerce électronique. *Systèmes d'information et gestion du commerce électronique*, 21(3), 527-570.
87. Hammouda, M. (2009). *Utilisation des techniques de data mining pour la modélisation du parcours scolaire et la prédiction du succès et du risque d'échec* (Doctoral dissertation, Blida).
88. Frimousse, S., & Peretti, J. M. (2019). «Expérience collaborateur» et «Expérience client»: comment l'entreprise peut-elle utiliser l'Intelligence Artificielle pour progresser?. *Question (s) de management*, (1), 135-156.
89. KATRAGADDA, V. (2022). Segmentation dynamique de la clientèle : utilisation de l'apprentissage automatique pour identifier et répondre aux divers besoins des clients en temps réel. *IRE Journals*, 5(10), 278-279.
90. Collica, R. S. (2017). *Customer segmentation and clustering using SAS Enterprise Miner*. Sas Institute.
91. Pranjic, G., & Rekettye, G. (2019). Interaction of the social media and big data in reaching marketing success in the era of the fourth industrial revolution. *International Journal of Business Performance Management*, 20(3), 247-260.
92. KATRAGADDA, V. (2022). Dynamic Customer Segmentation: Using Machine Learning to Identify and Address Diverse Customer Needs in Real-Time. *IRE Journals*, 5(10), 278-279.
93. Ranjan, A., et Srivastava, S. (2022, novembre). Segmentation de la clientèle à l'aide de l'apprentissage automatique : une revue de la littérature. Dans les *actes de la conférence AIP* (vol. 2481, n° 1). Éditions AIP.
94. Kasem, M. S., Hamada, M. et Taj-Eddin, I. (2024). Profilage des clients, segmentation et prédiction des ventes à l'aide de l'IA dans le marketing direct. *Informatique neuronale et applications*, 36(9), 4995-5005.
95. Dullaghan, C., & Rozaki, E. (2017). Integration of machine learning techniques to evaluate dynamic customer segmentation analysis for mobile customers. *arXiv preprint arXiv:1702.02215*.



Webographie

- w1. [Données sur les clients : Qu'est-ce que c'est, types et méthodes de collecte. \(questionpro.com\)](#) consulte le 28/02/2024
- w2. [RGPD : de quoi parle-t-on ? | CNIL](#) consulte le 28/02/2024
- w3. [Information géographique : définition et explications \(aquaportail.com\)](#) consulte le 28/02/2024
- w4. [What is Transactional Data? | TIBCO](#) consulte le 28/02/2024
- w5. [Data source : Définition et détails sur leur fonctionnement \(datascientest.com\)](#) consulte le 28/02/2024
- w6. [Source de données : définition et types | Talend](#) consulte le 28/02/2024
- w7. [Qu'est-ce que la segmentation de la clientèle ? Définition, modèles, analyse, stratégie et exemples \(ideascale.com\)](#) consulte le 28/02/2024
- w8. [Machine Learning : Définition, fonctionnement, utilisations \(datascientest.com\)](#) consulté le 28/02/2024
- w9. [Qu'est-ce que l'apprentissage supervisé ? | IBM](#) consulte le 28/02/2024
- w10. [Machine learning : quels sont les trois types d'apprentissage automatique ? \(cmim.fr\)](#) consulte le 28/02/2024
- w11. [Les principales méthodes d'apprentissage en Intelligence Artificielle \(IA\) \(pandia.pro\)](#) consulte le 28/02/2024
- w12. [Apprentissage semi-supervisé — DataFranca](#) consulté le 28/02/2024
- w13. [Algorithme Q-Learning : apprentissage par renforcement \(jedha.co\)](#) consulté le 10/12/2023
- w14. [Comment le machine learning améliore l'analyse de la segmentation client - FasterCapital](#) consulte le 28/02/2024
- w15. [L'impact de l'IA dans la conception d'interface - OCTO Talks !](#) consulte le 01/03/2024
- w16. [Design génératif : révolution ou évolution ? - Arquen](#) consulte le 01/03/2024
- w17. [Midjourney](#) consulte le 01/03/2024
- w18. [Experience Cloud AI Services with Adobe Sensei](#) consulte le 01/03/2024
- w19. [Framer — The internet is your canvas](#) consulte le 01/03/2024
- w20. [Fontjoy - Generate font pairings in one click](#) consulte le 01/03/2024
- w21. [Logo Design & Brand Identity for Entrepreneurs | Looka](#) consulte le 01/03/2024
- w22. [The home of exceptional digital experiences - Optimizely](#) consulte le 01/03/2024
- w23. [Hotjar : heatmaps de sites et outils d'analyse comportementale](#) consulte le 01/03/2024
- w24. [VisualEyes - Optimize your UX via Attention Heatmaps and Clarity](#) consulte le 01/03/2024
- w25. [Définition | Deep Learning - Apprentissage profond \(futura-sciences.com\)](#) consulte le 01/03/2024



- w26. [Deep learning : définition, fonctionnement et applications \(hubspot.fr\)](#) consulte le 01/03/2024
- w27. [Deep Learning et machine learning : différence entre les technologies de données – AWS \(amazon.com\)](#) consulte le 01/03/2024
- w28. [Deep learning vs Machine learning : quelle est la différence ? - IONOS](#) consulte le 01/03/2024
- w29. <https://octolis.com/fr/blog/segmentation-rfm> consulte le 09/03/2024
- w30. <https://www.wizville.com/ressources/lexique/segmentation-client#:~:text=Chapitre%201-,%20Qu'est%20ce%20que%20la%20segmentation%20client%20%3F,ou%20le%20comportement%20d'achat.> consulte le 09/03/2024
- w31. <https://fastercapital.com/fr/contenu/Segmentation-des-canaux---l-importance-de-la-segmentation-des-canaux-dans-le-marketing.html#Introduction---la-segmentation-des-canaux-dans-le-marketing> consulte le 09/03/2024
- w32. <https://fastercapital.com/fr/contenu/Segmentation-basee-sur-la-technologie---Segmentation-basee-sur-la-technologie---Atteindre-vos-clients-experts-en-technologie.html#qu-est-ce-que-c-est-et-comment-cela-fonctionne> consulte le 09/03/2024
- w33. [Que signifie Qualité des données? - Definition IT de LeMagIT](#) consulte le 12/03/2024
- w34. [IBM Watson Studio](#) consulte le 11/03/2024
- w35. [Qu'est-ce que le machine learning ? | Machine Learning | Google for Developers](#) consulte le 13/03/2024
- w36. [Téléchargement gratuit | Anaconda](#) consulte le 13/03/2024
- w37. [Le tutoriel Python — Documentation Python 3.12.2](#) consulte le 13/03/2024
- w38. [Valeurs séparées par des virgules — Wikipédia \(wikipedia.org\)](#) consulte le 13/03/2024
- w39. [K-Means Clustering en Python : un guide pratique – Real Python](#) consulte le 13/03/2024
- w40. [Algorithme de machine vectorielle de support \(SVM\) - GeeksforGeeks](#) consulte le 13/03/2024
- w41. [Qu'est-ce que Scikit-Learn : la bibliothèque de machine learning incontournable – Kanaries](#) consulte le 13/03/2024
- w42. [Projet Jupyter | Domicile](#) consulte le 14/03/2024
- w43. [Installation — Documentation Anaconda](#) consulte le 14/03/2024
- w44. [Qu'est-ce que le deep learning ? | Oracle France](#) consulte le 12/06/2024
- w45. [What are the Types of Machine Learning? - in Detail - AnalyticsLearn](#) consulte le 08/08/2024





Table des matières

Dédicace	i
Remerciements	ii
Résumé	iii
Sommaire.....	v
Liste des figures.....	vi
Liste des tableaux	viii
Sigles et abréviations.....	ix
Introduction générale.....	1
Chapitre 1 : Généralité sur la segmentation des clients et le machine Learning	3
1.1. Données clients.....	4
1.1.1. Définitions	4
1.1.2. Types de données	4
1.1.3. Les sources de données clients	5
1.2. La segmentation des clients.....	5
1.2.1. Définition.....	5
1.2.2. Les rôles de la segmentation des clients.....	6
1.2.3. Les méthodes de la segmentation des clients	7
1.3. Le machine learning	9
1.3.1. Définitions et fonctionnement	9
1.3.2. Les formes de machine learning.....	10
1.3.2.1. L'apprentissage supervisé	10
1.3.2.2. L'apprentissage non supervisé	11
1.3.2.3. L'apprentissage semi-supervisé	12
1.3.2.4. L'apprentissage par renforcement	13
1.3.2.5. Apprentissage profond ou Deep Learning.....	13
1.3.3. Les algorithmes de machine learning selon la forme	14
Chapitre 2 : Le Machine Learning et la Segmentation des clients	22
2.1. Les apports du machine learning dans la segmentation des clients	23
2.1.1. Amélioration de la précision dans les segments.....	24
2.1.2. Segmentation dynamique	24
2.1.3. Analyse prédictive	25
2.1.4. La personnalisation des offres.....	25



2.2.	Les enjeux.....	25
2.2.1.	La Tarification dynamique	25
2.2.2.	La Publicité ciblée	26
2.2.3.	La Recommandations de produits	26
2.2.4.	Les Campagnes marketing personnalisées	26
2.2.5.	La Fidélisation améliorée des clients	27
2.2.6.	La Prise de décision en temps réel	27
2.3.	Les contraintes liées à la mise en œuvre	27
2.3.1.	Accès et Disponibilité des données	27
2.3.2.	Qualité des données.....	28
2.3.3.	Manque d'expertises en ML et en data science.....	28
2.3.4.	Aspects légaux de collecte et d'utilisations des données personnelles	29
Chapitre 3 : Etat de l'art et positionnement.....		30
3.1.	Les algorithmes de ML pour la segmentation des clients	16
3.1.1.	Les algorithmes de sélections.....	16
3.1.2.	Les algorithmes de clustering.....	17
3.1.3.	Les algorithmes de classification	18
3.1.4.	Les modèles prédictifs.....	19
3.1.5.	L'apprentissage continu sur les données clients.....	19
3.2.	Le machine learning pour la segmentation des clients.....	31
3.2.1.	Outils d'analyse de données avec le ML.....	31
3.2.2.	Outils d'analyse des sentiments avec le ML	33
3.3.	Positionnement	36
Chapitre 4 : Etude de cas		36
4.1.	Cahier des charges.....	41
4.1.1.	Expression des besoins	41
4.1.2.	Présentation du corpus d'étude.....	42
4.1.3.	Choix des outils de travail	42
4.1.4.	Installation des outils et bibliothèques	43
4.2.	Implémentation.....	44
4.2.1.	Nettoyage des données	44
4.3.	Interprétation des résultats.....	44
Conclusion et Perspectives		58
Références		59



Articles scientifiques & livres	59
Webographie	67
Table des matières	70
Annexe 1 : Segmentation données clients	73
Annexe 2 : prédiction données clients.....	75
Annexe 3 : Segmentation données transactionnelles	79
Annexe 4 : prédiction données transactionnelles	81



Annexe 1 : Segmentation données clients

```
import pandas as pd
import plotly.express as px
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Chargement du fichier CSV
df = pd.read_csv('C:\\Users\\Serigne Bii\\Desktop\\dataset\\Dataset_Mor\\Infos_clients.csv', delimiter=';',
encoding='utf-8')

# Renommer la colonne 'État' en 'ETAT'
df.rename(columns={'État': 'ETAT'}, inplace=True)

# Sélection des 100 premières valeurs des colonnes 'Âge', 'Genre' et 'ETAT' et suppression des valeurs
manquantes
df_segmented = df[['Âge', 'Genre', 'ETAT']].dropna().head(100)

# Encodage des données catégoriques (Genre et ETAT) en valeurs numériques
df_segmented['Genre_encoded'] = df_segmented['Genre'].astype('category').cat.codes
df_segmented['ETAT_encoded'] = df_segmented['ETAT'].astype('category').cat.codes

# Normalisation des données
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_segmented[['Âge', 'Genre_encoded', 'ETAT_encoded']])

# Déterminer le nombre de clusters (exemple : 3 clusters)
num_clusters = 3 # Ajuster le nombre de clusters selon les besoins

# Application de l'algorithme K-means avec n_init défini explicitement
kmeans = KMeans(n_clusters=num_clusters, random_state=42, n_init=10) # n_init doit être un entier
df_segmented['Segment'] = kmeans.fit_predict(df_scaled)

# Préparer les données pour le diagramme en barres 2D
# Compter le nombre de clients par Âge, Genre, État et Segment
df_bar = df_segmented.groupby(['Âge', 'Genre', 'ETAT', 'Segment']).size().reset_index(name='Nombre de
Clients')

# Création du diagramme en barres 2D
fig_bar = px.bar(
    df_bar,
    x='Âge',
    y='Nombre de Clients',
    color='Segment',
    facet_col='Genre',
    facet_row='ETAT',
    barmode='group', # Barres groupées par Âge pour chaque Genre et État
    title='Segmentation des Clients par Âge, Genre et État avec K-means ',
    labels={'Âge': 'Âge', 'Nombre de Clients': 'Nombre de Clients', 'Segment': 'Segment', 'Genre': 'Genre',
'ETAT': 'État'},
    color_discrete_sequence=px.colors.qualitative.Plotly # Utilisation d'une palette de couleurs qualitative
)

# Ajouter les valeurs des barres sur le graphique
fig_bar.update_traces(texttemplate='% {text}', textposition='outside')

# Ajouter le titre et ajuster les labels
```



```
fig_bar.update_layout(  
    xaxis_title='Âge',  
    yaxis_title='Nombre de Clients',  
    coloraxis_colorbar=dict(  
        title="Segment",  
    ),  
    width=1000, # Largeur du graphique  
    height=800 # Hauteur du graphique  
)  
  
# Affichage du graphique  
fig_bar.show()
```



Annexe 2 : prédiction données clients

```
import pandas as pd
import plotly.graph_objects as go
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LinearRegression
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import classification_report, mean_squared_error, r2_score

# Chargement des données depuis le fichier CSV
df = pd.read_csv('C:\\Users\\Serigne Bii\\Desktop\\dataset\\Dataset_Mor\\Infos_clients.csv', delimiter=';',
encoding='utf-8')

# Renommer la colonne 'État' en 'ETAT'
df.rename(columns={'État': 'ETAT'}, inplace=True)

# Préparer les données
df_encoded = df[['Genre', 'Âge', 'ETAT']].dropna() # On retire les valeurs manquantes

# Encodage des variables catégoriques
le_genre = LabelEncoder()
le_etat = LabelEncoder()
df_encoded['Genre'] = le_genre.fit_transform(df_encoded['Genre'])
df_encoded['ETAT'] = le_etat.fit_transform(df_encoded['ETAT'])

# Définir les caractéristiques et les cibles pour les modèles de prédiction
X = df_encoded[['Âge', 'Genre']]
y_etat = df_encoded['ETAT']
y_genre = df_encoded['Genre']

# Division en ensembles d'entraînement et de test (utiliser toutes les données)
X_train = X
y_etat_train = y_etat
y_genre_train = y_genre

# Utiliser les mêmes données pour l'entraînement et les prédictions
X_test = X
y_etat_test = y_etat
y_genre_test = y_genre

# Modèle pour prédire l'Âge avec la régression linéaire
reg_age = LinearRegression()
reg_age.fit(X_train, df_encoded['Âge'])
y_age_pred = reg_age.predict(X_test)

# Modèle pour prédire l'État avec MLPClassifier
clf_etat = MLPClassifier(hidden_layer_sizes=(10,), max_iter=1000, random_state=42)
clf_etat.fit(X_train, y_etat_train)
y_etat_pred = clf_etat.predict(X_test)

# Modèle pour prédire le Genre avec un réseau de neurones (MLPClassifier)
model_genre = MLPClassifier(hidden_layer_sizes=(10,), max_iter=1000, random_state=42)
model_genre.fit(X_train, y_genre_train)
y_genre_pred = model_genre.predict(X_test)

# Prédiction de l'âge, de l'état et du genre pour le prochain client (le premier client de l'ensemble de test
comme exemple)
```



```
next_client_data = X_test.iloc[0:1] # Utiliser un DataFrame pour maintenir les noms de colonnes

next_client_age = reg_age.predict(next_client_data)[0]
next_client_pred_etat = clf_etat.predict(next_client_data)[0]
next_client_pred_etat_label = le_etat.inverse_transform([next_client_pred_etat])[0]
next_client_pred_genre = model_genre.predict(next_client_data)[0]
next_client_pred_genre_label = le_genre.inverse_transform([next_client_pred_genre])[0]

# Affichage des informations du prochain client
print(f"\nÂge du Prochain Client : {next_client_age:.2f}")
print(f"Genre du Prochain Client : {next_client_pred_genre_label}")
print(f"État Prédit pour le Prochain Client : {next_client_pred_etat_label}")

# Créer un DataFrame pour la visualisation 3D
df_3d_etat = pd.DataFrame({
    'Âge': X_test['Âge'],
    'Genre': le_genre.inverse_transform(X_test['Genre']),
    'État Réel': le_etat.inverse_transform(y_etat_test),
    'État Prédit': le_etat.inverse_transform(y_etat_pred)
})

# Ajouter les données du prochain client pour la visualisation
df_3d_etat = pd.concat([df_3d_etat, pd.DataFrame({
    'Âge': [next_client_age],
    'Genre': [next_client_pred_genre_label],
    'État Réel': [None], # Changer la valeur ici pour ne pas perturber l'encodage
    'État Prédit': [next_client_pred_etat_label]
})], ignore_index=True)

# Encodage des variables pour la visualisation 3D
df_3d_etat['Genre'] = le_genre.transform(df_3d_etat['Genre'])
df_3d_etat['État Réel'] = df_3d_etat['État Réel'].apply(lambda x: le_etat.transform([x])[0] if x in
le_etat.classes_ else -1)
df_3d_etat['État Prédit'] = le_etat.transform(df_3d_etat['État Prédit'])

# Graphique 3D de la prédiction de l'État et du Genre
fig_etat_sexe = go.Figure()

# Définir les couleurs pour chaque état
colors = {i: color for i, color in enumerate(['red', 'green', 'blue', 'purple'])} # Ajuster en fonction des
étiquettes de l'état

# Ajouter les points pour chaque état réel
for etat, color in colors.items():
    df_subset = df_3d_etat[df_3d_etat['État Réel'] == etat]
    fig_etat_sexe.add_trace(go.Scatter3d(
        x=df_subset['Âge'],
        y=df_subset['Genre'],
        z=df_subset['État Réel'],
        mode='markers',
        marker=dict(
            size=8,
            color=color, # Couleur en fonction de l'état réel
            opacity=0.7,
        ),
        name=le_etat.inverse_transform([etat])[0] # Nom de l'état pour la légende
    ))
```



```
# Ajouter le point de prédiction pour le prochain client
fig_etat_sexe.add_trace(go.Scatter3d(
    x=[next_client_age],
    y=[le_genre.transform([next_client_pred_genre_label])[0]],
    z=[le_etat.transform([next_client_pred_etat_label])[0]],
    mode='markers+text',
    marker=dict(
        color='black',
        size=12,
        symbol='cross'
    ),
    text=[f'{next_client_pred_etat_label}\n{next_client_pred_genre_label}'],
    textposition='top center',
    name='Prochain Client Prédit'
))

# Mise en forme du graphique
fig_etat_sexe.update_layout(
    scene=dict(
        xaxis_title='Âge',
        yaxis_title='Genre',
        zaxis_title='État',
        xaxis=dict(
            title='Âge'
        ),
        yaxis=dict(
            tickvals=[0, 1], # Les genres sont placés sur l'axe Y comme 0 pour autre, 1 pour Male et 2 pour
            Female
            ticktext=le_genre.classes_, # Affiche les noms des genres
            title='Genre'
        ),
        zaxis=dict(
            tickvals=[0, 1, 2, 3], # Ajuster en fonction des étiquettes de l'état
            ticktext=le_etat.classes_, # Affiche les états
            title='État'
        ),
        aspectmode='cube' # Aspect cube pour une meilleure perspective
    ),
    title='Prédiction de l'âge, État et Genre du Prochain Client',
    width=900,
    height=700,
    showlegend=True
)

# Affichage du graphique
fig_etat_sexe.show()

# Calcul des métriques de performance
mse_age = mean_squared_error(y_etat, y_age_pred)
r2_age = r2_score(y_etat, y_age_pred)

# Obtenir les labels des classes réelles et prédites
labels_etat = le_etat.classes_
labels_genre = le_genre.classes_
```



```
# Calculer les rapports de classification en spécifiant les labels et zero_division pour éviter les avertissements
classification_report_etat = classification_report(y_etat_test, y_etat_pred,
labels=le_etat.transform(labels_etat), target_names=labels_etat, zero_division=0)
classification_report_genre = classification_report(y_genre_test, y_genre_pred,
labels=le_genre.transform(labels_genre), target_names=labels_genre, zero_division=0)

print(f"\nMétriques pour la Prédiction de l'Âge:")
print(f"Mean Squared Error (MSE): {mse_age:.2f}")
print(f"R^2 Score: {r2_age:.2f}\n")

print(f"\nClassification Report pour l'État:")
print(classification_report_etat)

print(f"\nClassification Report pour le Genre:")
print(classification_report_genre)
```



Annexe 3 : Segmentation données transactionnelles

```
import pandas as pd
import plotly.graph_objects as go
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import plotly.express as px

# Chargement des données depuis le fichier CSV
df = pd.read_csv('C:\\Users\\Serigne Bii\\Desktop\\dataset\\Dataset_Mor\\Transactions.csv', delimiter=';',
encoding='utf-8')

# Nettoyage des noms de colonnes
df.columns = df.columns.str.strip()

# Nettoyage des données pour 'Prix de Vente'
df['Prix de Vente'] = df['Prix de Vente'].str.replace('[\$,]', '', regex=True).str.replace('[\u202f]', '',
regex=True).astype(float)

# Remplacement des valeurs 'FAUX' et 'VRAI'
df['Commande en ligne'] = df['Commande en ligne'].replace({'FAUX': 'Non Commande en Ligne', 'VRAI':
'Commande en Ligne'})

# Sélection des colonnes pertinentes et récupération des 100 premières lignes
df_segmented = df[['Commande en ligne', 'Marque', 'Prix de Vente']].head(100)

# Normalisation des données pour 'Prix de Vente'
scaler = StandardScaler()
df_segmented['Prix de Vente_scaled'] = scaler.fit_transform(df_segmented[['Prix de Vente']])

# Application de l'algorithme K-means avec 3 clusters
kmeans = KMeans(n_clusters=len(df_segmented['Marque'].unique()), random_state=42, n_init=10)
df_segmented['Segment'] = kmeans.fit_predict(df_segmented[['Prix de Vente_scaled']])

# Couleurs pour chaque marque
colors = px.colors.qualitative.Plotly
marque_color_map = {marque: colors[i % len(colors)] for i, marque in
enumerate(df_segmented['Marque'].unique())}

# Création du graphique 3D avec chaque marque ayant sa propre couleur et segment
fig = go.Figure()

# Ajout des points pour les 100 premières marques avec couleurs différentes et segments uniques
for marque in df_segmented['Marque'].unique():
    df_subset = df_segmented[df_segmented['Marque'] == marque]
    fig.add_trace(go.Scatter3d(
        x=df_subset['Commande en ligne'],
        y=df_subset['Prix de Vente'],
        z=df_subset['Segment'],
        mode='markers',
        marker=dict(
            size=5,
            opacity=0.8,
            color=marque_color_map[marque], # Couleur de la marque
        ),
        name=marque
    ))
```



```
# Mise en forme du graphique
fig.update_layout(
    scene=dict(
        xaxis_title='Commande en Ligne',
        yaxis_title='Prix de Vente',
        zaxis_title='Segment',
        aspectmode='cube' # Aspect cube pour une meilleure perspective
    ),
    title='Segmentation des Transactions par Commande en Ligne, Marque et Prix de Vente avec K-means',
    width=800,
    height=600,
    showlegend=True
)

# Affichage du graphique
fig.show()
```



Annexe 4 : prédiction données transactionnelles

```
import pandas as pd
import plotly.graph_objects as go
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import GradientBoostingClassifier, RandomForestRegressor
from sklearn.metrics import precision_score, recall_score, f1_score, mean_absolute_error,
mean_squared_error
import numpy as np

# 1. Chargement et Prétraitement des Données
df = pd.read_csv('C:\\Users\\Serigne Bii\\Desktop\\dataset\\Dataset_Mor\\Transactions.csv', delimiter=';',
encoding='utf-8')
df.columns = df.columns.str.strip() # Nettoyage des noms de colonnes
df['Prix de Vente'] = df['Prix de Vente'].str.replace('\\$', '', regex=True).str.replace('\\u202f', '',
regex=True).astype(float)

# 2. Sélection et Préparation des Données
df_segmented = df[['Commande en ligne', 'Marque', 'Prix de Vente']].head(100).copy()
df_segmented['ClientID'] = df_segmented.groupby('Marque').ngroup()
df_segmented['Next_Marque'] = df_segmented.groupby('ClientID')['Marque'].shift(-1)
df_segmented['Next_Commande_en_Ligne'] = df_segmented.groupby('ClientID')['Commande en
ligne'].shift(-1)
df_segmented['Next_Prix_de_Vente'] = df_segmented.groupby('ClientID')['Prix de Vente'].shift(-1)
df_segmented.dropna(subset=['Next_Marque', 'Next_Commande_en_Ligne', 'Next_Prix_de_Vente'],
inplace=True)

# 3. Encodage et Préparation des Variables
encoder = LabelEncoder()
df_segmented['Marque'] = encoder.fit_transform(df_segmented['Marque'])
df_segmented['Next_Marque'] = df_segmented['Next_Marque'].apply(lambda x: x if x in encoder.classes_
else None)
df_segmented['Next_Marque'] = df_segmented['Next_Marque'].map(lambda x: encoder.transform([x])[0] if
pd.notna(x) else -1)
df_segmented['Commande en ligne'] = df_segmented['Commande en ligne'].map({'VRAI': 1, 'FAUX': 0})
df_segmented['Next_Commande_en_Ligne'] = df_segmented['Next_Commande_en_Ligne'].map({'VRAI':
1, 'FAUX': 0})

X = df_segmented[['Marque', 'Commande en ligne', 'Prix de Vente']]
y_marque = df_segmented['Next_Marque']
y_commande = df_segmented['Next_Commande_en_Ligne']
y_prix = df_segmented['Next_Prix_de_Vente']

# 4. Entraînement des Modèles
X_train, X_test, y_marque_train, y_marque_test, y_commande_train, y_commande_test, y_prix_train,
y_prix_test = train_test_split(
    X, y_marque, y_commande, y_prix, test_size=0.3, random_state=42
)

model_marque = GradientBoostingClassifier(n_estimators=100, random_state=42)
model_marque.fit(X_train, y_marque_train)

model_commande = GradientBoostingClassifier(n_estimators=100, random_state=42)
model_commande.fit(X_train, y_commande_train)

model_prix = RandomForestRegressor(n_estimators=100, random_state=42)
```



```
model_prix.fit(X_train, y_prix_train)

# 5. Évaluation des Modèles
y_marque_pred = model_marque.predict(X_test)
y_commande_pred = model_commande.predict(X_test)
y_prix_pred = model_prix.predict(X_test)

precision_marque = precision_score(y_marque_test, y_marque_pred, average='weighted')
recall_marque = recall_score(y_marque_test, y_marque_pred, average='weighted')
f1_marque = f1_score(y_marque_test, y_marque_pred, average='weighted')

precision_commande = precision_score(y_commande_test, y_commande_pred, average='weighted')
recall_commande = recall_score(y_commande_test, y_commande_pred, average='weighted')
f1_commande = f1_score(y_commande_test, y_commande_pred, average='weighted')

mae_prix = mean_absolute_error(y_prix_test, y_prix_pred)
mse_prix = mean_squared_error(y_prix_test, y_prix_pred)
rmse_prix = np.sqrt(mse_prix)

print("\nMétriques de Performance :")
print(f"Précision (Marque) : {precision_marque:.2f}")
print(f"Rappel (Marque) : {recall_marque:.2f}")
print(f"Score F1 (Marque) : {f1_marque:.2f}")
print(f"Précision (Commande en Ligne) : {precision_commande:.2f}")
print(f"Rappel (Commande en Ligne) : {recall_commande:.2f}")
print(f"Score F1 (Commande en Ligne) : {f1_commande:.2f}")
print(f"MAE (Prix de Vente) : {mae_prix:.2f}")
print(f"MSE (Prix de Vente) : {mse_prix:.2f}")
print(f"RMSE (Prix de Vente) : {rmse_prix:.2f}")

# 6. Prédiction et Visualisation
next_marque_pred = model_marque.predict(X_test.iloc[[0]])[0]
next_commande_pred = model_commande.predict(X_test.iloc[[0]])[0]
next_prix_pred = model_prix.predict(X_test.iloc[[0]])[0]

next_marque_pred_label = None
if next_marque_pred != -1:
    try:
        next_marque_pred_label = encoder.inverse_transform([next_marque_pred])[0]
    except ValueError:
        next_marque_pred_label = "Marque inconnue"
else:
    next_marque_pred_label = "Marque inconnue"

print("\nProchaine Transaction Prédite :")
print(f"Marque : {next_marque_pred_label}")
print(f"Commande en Ligne : {'Oui' if next_commande_pred == 1 else 'Non'}")
print(f"Prix de Vente : {next_prix_pred:.2f}")

df_pred = pd.DataFrame({
    'Marque Prédite': [next_marque_pred_label],
    'Commande en Ligne Prédite': ['Oui' if next_commande_pred == 1 else 'Non'],
    'Prix de Vente Prédit': [next_prix_pred]
})

print("\nRésumé de la Prochaine Transaction Prédite :")
print(df_pred)
```



```
df_visualize = df_segmented[['Marque', 'Commande en ligne', 'Prix de Vente']].copy()
df_visualize['Marque'] = encoder.inverse_transform(df_visualize['Marque'])
df_visualize['Commande en ligne'] = df_segmented['Commande en ligne'].map({0: 'Non Commande en Ligne', 1: 'Commande en Ligne'})

unique_marques = df_visualize['Marque'].unique()
colors = ['blue', 'orange', 'green', 'red', 'purple', 'cyan', 'magenta'] # Ajouter plus de couleurs si nécessaire

fig = go.Figure()

for i, marque in enumerate(unique_marques):
    df_subset = df_visualize[df_visualize['Marque'] == marque]
    fig.add_trace(go.Scatter3d(
        x=df_subset['Marque'],
        y=df_subset['Prix de Vente'],
        z=[i] * len(df_subset),
        mode='markers',
        marker=dict(
            size=8,
            opacity=0.8,
            color=colors[i % len(colors)],
            symbol='circle',
        ),
        name=marque
    ))

if next_marque_pred_label != "Marque inconnue":
    next_marque_encoded = encoder.transform([next_marque_pred_label])[0] if next_marque_pred_label in encoder.classes_ else None
else:
    next_marque_encoded = None

fig.add_trace(go.Scatter3d(
    x=[next_marque_encoded] if next_marque_encoded is not None else [None],
    y=[next_prix_pred],
    z=[next_marque_encoded] if next_marque_encoded is not None else [None],
    mode='markers+text',
    marker=dict(
        color='black',
        size=12,
        opacity=1.0,
        symbol='cross'
    ),
    text=[f'{next_marque_pred_label}\n{"Commande en Ligne" if next_commande_pred == 1 else "Non Commande en Ligne"}\n{next_prix_pred:.2f}'],
    textposition='top center',
    name='Prochaine Transaction Prédite'
))

fig.update_layout(
    scene=dict(
        xaxis_title='Marque',
        yaxis_title='Prix de Vente',
        zaxis_title='Segment',
        xaxis=dict(
            tickvals=[encoder.transform([m])[0] for m in encoder.classes_],
```



```
ticktext=encoder.classes_,
title='Marque'
),
zaxis=dict(
title='Segment',
range=[-1, len(unique_marques)]
),
aspectmode='cube'
),
title='Marques Actuelles et Prochaine Transaction Prédite',
width=900,
height=700,
showlegend=True
)
fig.show()
```