**Objectif en termes commerciaux dans l'analyse du churn**

Contexte :

Le churn, ou taux d'attrition, désigne la perte de clients au fil du temps. Dans un contexte commercial, il s'agit d'un indicateur clé de performance (KPI) qui mesure la fidélité des clients et la capacité d’une entreprise à les retenir.

Prédire et réduire le churn pour maximiser la rétention client et la rentabilité à long terme.

**Déclinaisons des objectifs commerciaux :**

1. Détection précoce des clients à risque

* Identifier les clients ayant une probabilité élevée d’abandonner (via la colonne Churn\_Probability dans le dataset).
* Fournir des scores de risque pour hiérarchiser les actions à mener.
* Minimiser les pertes de revenus associées au départ des clients.

2. Optimisation des stratégies de fidélisation

* Personnalisation : Adapter les programmes de fidélité, les promotions et les campagnes marketing en fonction des comportements identifiés.
* Prévention ciblée : Offrir des remises ou avantages spécifiques aux clients à haut risque pour les retenir.
* Effort marketing proactif : Déployer des campagnes d’engagement en fonction des préférences temporelles (colonne Preferred\_Purchase\_Times) et géographiques (Region).

3. Segmentation client

* Analyser les groupes de clients selon leur région, leurs habitudes d’achat et leurs valeurs de vie client (Lifetime\_Value) afin de concentrer les ressources marketing là où elles sont les plus efficaces.
* Créer des profils types pour cibler différents segments avec des approches spécifiques.

4. Prévision des revenus futurs

* Estimer l’impact financier d’un churn élevé.
* Projeter les pertes potentielles sur la base des valeurs de vie client (Lifetime\_Value) et ajuster les budgets marketing en conséquence.
* Quantifier l’amélioration attendue après des actions correctives.

5. Augmentation de la satisfaction client

* Identifier les motifs de désengagement en analysant les variables comportementales :
  + Temps entre les achats (Time\_Between\_Purchases)
  + Valeur des commandes (Average\_Order\_Value)
* Améliorer l’expérience utilisateur et le service client grâce à ces informations.

**Indicateurs de performance (KPIs) associés :**

* Taux de churn (%) = (Nombre de clients perdus / Nombre total de clients) × 100
* Taux de rétention (%) = 100 - Taux de churn
* Valeur vie client moyenne (CLV) = Somme des revenus générés par un client pendant sa durée de vie.
* Taux de conversion des stratégies de fidélisation (%) : Mesure l'efficacité des campagnes.
* ROI des actions de rétention : Bénéfices générés vs coûts des stratégies déployées.

**Utilisation de la solution pour la problématique du churn**

**1. Analyse prédictive pour identifier les clients à risque**

Fonctionnalité :

* Construire un modèle de machine learning capable d’estimer la probabilité de churn pour chaque client en utilisant des variables clés comme :
  + Comportement d'achat (ex. : fréquence d'achat, délais entre les commandes).
  + Valeur client (Lifetime\_Value, Average\_Order\_Value).
  + Données démographiques et géographiques (région).
  + Stratégies de rétention appliquées.

Utilisation :

* Générer une liste priorisée des clients les plus susceptibles d’abandonner.
* Identifier les tendances et caractéristiques des clients à haut risque pour des interventions ciblées.

**2. Segmentation des clients pour des actions marketing personnalisées**

Fonctionnalité :

* Catégoriser les clients en segments distincts selon leurs comportements d’achat et leurs préférences :
  + Acheteurs occasionnels vs acheteurs fréquents.
  + Clients premium (forte valeur vie) vs clients à faible contribution.
  + Préférences temporelles (matin, soir) et saisonnalité (été, hiver).

Utilisation :

* Adapter les offres promotionnelles et les stratégies de fidélisation pour chaque segment.
* Par exemple :
  + Offrir des réductions aux clients à risque identifiés via le modèle.
  + Renforcer les programmes de fidélité pour les clients fidèles.

**3. Surveillance continue et reporting en temps réel**

Fonctionnalité :

* Mettre en place des tableaux de bord interactifs pour suivre :
  + L’évolution du taux de churn en temps réel.
  + Les performances des campagnes de rétention.
  + Les revenus récupérés grâce aux actions mises en place.

Utilisation :

* Aider les équipes commerciales et marketing à ajuster rapidement leurs stratégies.
* Fournir des rapports périodiques aux décideurs pour mesurer l’efficacité des efforts de rétention.

**4. Simulation et optimisation des stratégies de fidélisation**

Fonctionnalité :

* Tester virtuellement différentes stratégies avant leur déploiement.
* Exemple : Simuler l’impact d’un programme de fidélité (réduction de 10 %) sur la probabilité de churn.

Utilisation :

* Prévoir les résultats financiers des campagnes avant leur mise en œuvre.
* Optimiser les budgets marketing pour maximiser le retour sur investissement (ROI).

**5. Intégration dans les workflows métier existants**

Fonctionnalité :

* Connecter la solution aux CRM et plateformes marketing (par exemple Salesforce, HubSpot).
* Automatiser les alertes pour signaler les clients à risque aux équipes commerciales.

Utilisation :

* Envoyer des emails automatiques avec des promotions personnalisées aux clients identifiés comme vulnérables.
* Générer des listes de rappels téléphoniques pour contacter directement les clients à haut risque.

**6. Amélioration continue grâce à l’apprentissage automatique**

Fonctionnalité :

* Mettre à jour le modèle au fil du temps avec de nouvelles données :
  + Nouveaux comportements d’achat.
  + Résultats des campagnes passées.

Utilisation :

* Assurer une précision accrue dans la détection du churn au fur et à mesure que des données supplémentaires sont collectées.
* Identifier les tendances émergentes avant qu’elles ne deviennent critiques.

**Solutions ou alternatives actuelles pour la gestion du churn**

**1. Méthodes analytiques traditionnelles**

Analyse descriptive et statistique

* Principe : Étudier les données historiques des clients pour comprendre les tendances et les causes potentielles du churn.
* Outils : Excel, SQL, Google Analytics.
* Avantages :
  + Simple à mettre en place.
  + Fournit des insights rapides sur les données de base.
* Inconvénients :
  + Limité à des analyses rétrospectives (ce qui s’est passé) au lieu de prévisions (ce qui va se passer).
  + Moins efficace pour détecter des relations complexes entre les variables.

Segmentation basée sur les règles métier

* Principe : Utiliser des règles prédéfinies (ex. : clients inactifs depuis 30 jours) pour identifier les clients à risque.
* Outils : CRM (Salesforce, HubSpot).
* Avantages :
  + Facile à comprendre et à expliquer aux équipes métiers.
  + Peut être intégrée rapidement dans des systèmes existants.
* Inconvénients :
  + Approche rigide qui ne capte pas bien les comportements complexes ou les signaux faibles.
  + Ne s’adapte pas automatiquement aux nouveaux modèles de comportement client.

**2. Solutions basées sur l’intelligence artificielle et le machine learning**

Modèles prédictifs supervisés

* Principe : Construire un modèle de machine learning (régression logistique, forêts aléatoires, XGBoost, etc.) entraîné sur les données historiques pour prédire la probabilité de churn.
* Outils : Python (Scikit-learn, TensorFlow, PyTorch), R, AutoML (H2O.ai, Google AutoML).
* Avantages :
  + Fournit des prédictions précises et adaptables aux nouvelles données.
  + Identifie les variables les plus importantes contribuant au churn.
* Inconvénients :
  + Nécessite des compétences techniques en science des données et en modélisation.
  + Peut être difficile à interpréter pour les utilisateurs métiers (problème de boîte noire).

Modèles non supervisés

* Principe : Utiliser des algorithmes de clustering (ex. : k-means, DBSCAN) pour regrouper les clients selon des caractéristiques similaires et détecter les groupes à haut risque.
* Avantages :
  + Ne nécessite pas d’étiquettes de churn préexistantes.
  + Utile pour l’exploration initiale et la segmentation avancée.
* Inconvénients :
  + Moins précis pour les prévisions spécifiques sur un client donné.
  + Nécessite des ajustements manuels pour interpréter les résultats.

**3. Solutions basées sur l’apprentissage profond (Deep Learning)**

* Principe : Utiliser des réseaux neuronaux (ANN, RNN, LSTM) pour identifier des schémas complexes dans les comportements d’achat.
* Outils : TensorFlow, Keras, PyTorch.
* Avantages :
  + Capable d’exploiter de grandes quantités de données non structurées (textes, journaux d’activité).
  + Très performant pour détecter des relations non linéaires et des tendances subtiles.
* Inconvénients :
  + Consommation de ressources importantes.
  + Nécessite des compétences avancées en IA.
  + Moins transparent et difficile à expliquer.

**4. Outils SaaS et plateformes commerciales**

* Principe : Utiliser des plateformes prêtes à l’emploi pour le suivi du churn et la gestion des relations clients.
* Exemples :
  + Tableau : Visualisation et tableaux de bord.
  + Salesforce Einstein : IA intégrée pour les prévisions.
  + Zendesk : Suivi des interactions client et gestion des tickets.
  + Mixpanel et Amplitude : Analyse des comportements d’utilisation des produits.
  + ChurnZero et ProfitWell : Spécialisés dans la prévision et la réduction du churn.
* Avantages :
  + Facile à intégrer dans les flux de travail existants.
  + Adapté aux utilisateurs non techniques.
* Inconvénients :
  + Coût élevé, surtout pour les PME.
  + Moins personnalisables que des solutions développées en interne.

**5. Approches heuristiques et marketing basées sur les retours clients**

* Principe : Collecter des avis, des commentaires et des enquêtes pour comprendre les raisons du churn.
* Exemples : Enquêtes Net Promoter Score (NPS), formulaires d’abandon de panier, ou entretiens.
* Avantages :
  + Permet d’explorer les causes qualitatives du churn (insatisfaction, concurrence).
  + Simple et rapide à mettre en œuvre.
* Inconvénients :
  + Données souvent biaisées ou incomplètes.
  + Moins utiles pour des analyses quantitatives et prédictives.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Solution** | **Précision** | **Facilité d’implémentation** | **Coût** | **Adaptabilité** |
| Analyse descriptive | Faible | Facile | Faible | Peu flexible |
| Segmentation manuelle | Moyenne | Moyenne | Faible | Moyenne |
| Machine learning supervisé | Élevée | Moyenne à difficile | Moyen | Très flexible |
| Deep learning | |  | | --- | |  |  |  | | --- | | Très élevée | | Difficile | Élevé | Très flexible |
| Outils SaaS | Moyenne | Facile | Élevé | Moyenne |
| Feedback client et heuristiques | Faible | Facile | Faible | Peu flexible |

**Définition du problème supervisé**

Le problème du churn dans ce contexte est un problème d’apprentissage supervisé. La variable cible Churn\_Probability sert à entraîner un modèle qui prédit la probabilité qu’un client se désengage. Ce problème peut être traité soit comme une régression pour prédire une probabilité continue, soit comme une classification binaire en définissant un seuil (par exemple 0,5) pour catégoriser les clients à risque.

Concernant le traitement, il peut être hors ligne (batch) pour analyser les données historiques et entraîner le modèle, ou en ligne (streaming) pour des prédictions en temps réel et des réactions immédiates face aux comportements des clients.

Enfin, bien qu’il soit supervisé, des approches non supervisées comme le clustering peuvent compléter l’analyse. Elles permettent de segmenter les clients ou de détecter des anomalies dans les comportements.

En résumé, il s’agit principalement d’un problème supervisé avec des méthodes de régression ou de classification, traité en mode batch ou streaming selon les besoins, et pouvant intégrer des outils non supervisés pour enrichir l’analyse.

**Mesurer la performance**

Pour mesurer la performance dans un problème de churn, plusieurs métriques sont utilisées en fonction de l’approche adoptée : régression ou classification.

Dans le cas d’une régression, où l’objectif est de prédire une probabilité continue, les mesures principales incluent :

* Erreur quadratique moyenne (MSE) et Erreur absolue moyenne (MAE) pour évaluer la précision des prédictions.
* R² (coefficient de détermination) pour mesurer dans quelle mesure les variations des prédictions expliquent les variations des valeurs réelles.

Pour une classification binaire, où l’on catégorise les clients en churn ou non-churn, les métriques sont :

* Précision : Proportion des prédictions correctes.
* Rappel (Recall) : Capacité à identifier les clients churn correctement.
* F1-Score : Moyenne harmonique entre précision et rappel, équilibrant les deux métriques.
* AUC-ROC (aire sous la courbe) : Indicateur global de performance en mesurant la capacité du modèle à distinguer entre churn et non-churn.

Dans un contexte commercial, il est aussi pertinent de suivre des indicateurs spécifiques comme :

* Coût d’acquisition client (CAC) pour vérifier si les stratégies mises en place sont rentables.
* Retour sur investissement (ROI) des campagnes de rétention.
* Taux de churn réel avant et après l’application des solutions pour évaluer leur efficacité.

Les mesures de performance combinent des métriques statistiques (AUC-ROC, F1-Score) avec des indicateurs métier (ROI, CAC) pour assurer à la fois la précision technique et la pertinence commerciale des résultats.

**Alignement de performance avec les objectifs business**

La mesure de la performance est alignée avec les objectifs business si les métriques utilisées reflètent directement l’impact des prédictions sur la rétention client et la rentabilité.

Dans le cas du churn, l’objectif commercial principal est de réduire la perte de clients et d’augmenter la fidélisation tout en optimisant les coûts d’acquisition et de rétention. Pour répondre à cet objectif, les métriques techniques comme l’AUC-ROC, le F1-Score ou la précision doivent être reliées à des indicateurs métiers tels que :

* Réduction du taux de churn réel mesuré avant et après la mise en place des actions correctives. Une diminution significative prouve l’efficacité de la solution.
* Amélioration de la valeur vie client (Lifetime Value) en suivant si les clients identifiés comme à risque et retenus génèrent davantage de revenus à long terme.
* Retour sur investissement (ROI) des campagnes marketing et des actions de rétention, permettant d’évaluer si les dépenses engagées pour fidéliser les clients sont compensées par les revenus supplémentaires générés.
* Taux de conversion des interventions pour vérifier combien de clients à risque ont changé de comportement après l’application des stratégies ciblées.

Ainsi, les métriques techniques assurent que le modèle est précis et fiable, tandis que les métriques commerciales valident son impact direct sur la rentabilité et la croissance. Pour garantir un alignement optimal, il est essentiel d’intégrer les deux approches dans un tableau de bord combinant des indicateurs prédictifs et des résultats observés.

**Performance minimale requise pour atteindre les objectifs business**

La performance minimale requise pour atteindre les objectifs business dépend de la capacité du modèle à réduire efficacement le churn tout en optimisant les coûts de rétention.

Sur le plan technique, une performance acceptable repose sur des métriques précises :

* AUC-ROC : Un score supérieur à 0.80 est généralement considéré comme bon, car il indique une forte capacité à distinguer les clients à risque de ceux qui ne le sont pas. En dessous de 0.70, le modèle serait jugé insuffisant pour des décisions critiques.
* F1-Score : Une valeur d’au moins 0.75 est souhaitable pour équilibrer précision et rappel, assurant ainsi que les clients à risque sont identifiés sans générer trop de faux positifs.
* Précision et rappel : Ces métriques doivent idéalement être au-dessus de 75 %, surtout pour minimiser les faux négatifs, car manquer un client à risque peut entraîner une perte de revenus directe.

Sur le plan commercial, les résultats doivent se traduire par des gains mesurables :

* Réduction du taux de churn : Une baisse d’au moins 10 % dans le taux d’attrition est souvent un objectif réaliste pour justifier les investissements dans l’IA. Par exemple, si le taux initial est de 20 %, l’objectif serait de le ramener à 18 % ou moins.
* Amélioration de la valeur vie client (CLV) : Une augmentation moyenne de 5 à 10 % par client grâce à des campagnes de fidélisation ciblées est un indicateur clé.
* Retour sur investissement (ROI) : Le coût des campagnes et de l’infrastructure IA doit être compensé par une hausse des revenus ou une réduction des coûts d’acquisition d’au moins 20 %.

La performance minimale requise combine des métriques techniques solides (AUC-ROC > 0.80 et F1 > 0.75) avec des impacts commerciaux clairs, tels qu’une baisse du churn d’au moins 10 % et un ROI positif. Ce seuil garantit à la fois la fiabilité du modèle et son utilité pour soutenir la stratégie de rétention.

**Problème existant & outils existants**

Le problème du churn partage de nombreuses similitudes avec d’autres problématiques liées à la prévision et à la classification dans divers secteurs. Ces problèmes comparables permettent de tirer parti d’expériences antérieures et d’outils préexistants pour accélérer le développement d’une solution adaptée.

Dans le domaine des abonnements et des services SaaS, un problème comparable est la prédiction des résiliations d’abonnement. Les entreprises analysent des comportements d’utilisation, comme la fréquence de connexion ou l’engagement avec les fonctionnalités, pour détecter les utilisateurs susceptibles de partir. Des outils comme HubSpot et Salesforce Einstein sont fréquemment utilisés pour automatiser ces prédictions grâce à des algorithmes supervisés tels que les forêts aléatoires et XGBoost. Ces approches peuvent être adaptées pour identifier les clients à risque dans un contexte de churn commercial.

Un autre parallèle peut être fait avec les systèmes de détection de fraude dans le secteur bancaire. Bien que la fraude repose souvent sur des anomalies ponctuelles et le churn sur des tendances comportementales, les deux reposent sur des modèles capables d’identifier des schémas anormaux. Les algorithmes tels que Isolation Forest ou les réseaux neuronaux récurrents (RNN) utilisés pour la détection de fraude peuvent être réutilisés pour identifier des comportements inhabituels menant au churn, en particulier si l’on analyse des séries temporelles comme les dates d’achat ou les périodes d’inactivité.

Dans le secteur des réseaux sociaux et des applications mobiles, la prévision de la désinscription d’utilisateurs suit une logique similaire. Les entreprises surveillent les interactions, les clics et la durée des sessions pour détecter des signes de désengagement. Des techniques telles que l’analyse de survie sont employées pour estimer le moment probable d’un départ. Cette méthodologie pourrait être appliquée pour analyser les périodes entre achats et détecter des tendances avant qu’un client ne se désengage complètement.

Les outils déjà existants pour ces cas d’usage, tels que Scikit-learn, TensorFlow ou des plateformes comme Azure Machine Learning, offrent des bibliothèques et des pipelines préconstruits pour le traitement et la modélisation des données. Par exemple, des modules comme SMOTE pour gérer les déséquilibres dans les données et GridSearchCV pour optimiser les hyperparamètres peuvent être directement appliqués au churn. De plus, des solutions comme H2O.ai et DataRobot permettent d’automatiser la sélection des modèles et des hyperparamètres tout en fournissant des tableaux de bord pour l’analyse des résultats.

Les problèmes liés au churn s’apparentent à la détection de fraude, aux prévisions de résiliations d’abonnements et aux analyses de désengagement d’utilisateurs. Ces expériences et outils préexistants peuvent être réutilisés pour gagner du temps et améliorer la performance du modèle. L’intégration d’approches éprouvées, combinée à des outils de machine learning avancés, offre une base solide pour aborder cette problématique efficacement.

**Expertise humaine du problème**

L’expertise humaine sur ce problème est bien disponible et peut être mobilisée à différents niveaux pour répondre aux exigences analytiques et commerciales liées au churn.

Sur le plan technique, les data scientists et analystes de données possèdent les compétences nécessaires pour préparer et traiter les données, entraîner des modèles prédictifs et évaluer leur performance. Ils maîtrisent des outils comme Scikit-learn, TensorFlow et H2O.ai, qui permettent de créer des solutions basées sur des algorithmes supervisés ou non supervisés. Cette expertise est essentielle pour garantir que les modèles sont robustes et capables de généraliser leurs prédictions sur de nouveaux jeux de données.

Du côté métier, les équipes marketing et commerciales apportent une compréhension approfondie des comportements clients et des stratégies de fidélisation. Leur rôle est d’interpréter les résultats des modèles et de traduire les prédictions en actions concrètes, comme des offres promotionnelles ciblées ou des relances par email. Leur expérience terrain est également précieuse pour valider les hypothèses utilisées dans les modèles et affiner les indicateurs de performance en fonction des objectifs business.

Enfin, les décideurs stratégiques jouent un rôle clé dans l’alignement des efforts techniques et commerciaux. Ils évaluent l’impact financier des solutions développées et s’assurent que les investissements réalisés en termes d’infrastructure et de campagnes marketing produisent un retour mesurable. Leur expertise est indispensable pour prioriser les actions en fonction des ressources disponibles et des opportunités de marché.

En combinant ces différentes expertises, il est possible de développer une solution complète qui intègre des outils analytiques avancés et des stratégies commerciales ciblées. Cela garantit une approche équilibrée entre précision technique et pertinence opérationnelle.

**Résoudre le problème manuellement**

Pour résoudre manuellement le problème du churn, il serait nécessaire d’adopter une approche analytique traditionnelle basée sur l’étude des comportements clients et des indicateurs clés sans utiliser d’algorithmes avancés.

La première étape consisterait à collecter et organiser les données clients afin d’identifier les variables pertinentes. On commencerait par examiner des caractéristiques comme la fréquence des achats, la valeur moyenne des commandes et le temps écoulé depuis le dernier achat. Ces indicateurs permettraient de détecter des schémas de comportement susceptibles d’indiquer un risque d’attrition, par exemple une baisse progressive des dépenses ou un allongement des délais entre deux transactions.

Ensuite, il serait possible d’établir des seuils critiques pour différencier les clients à risque des clients fidèles. Par exemple, un client qui n’a pas effectué d’achat au cours des 60 derniers jours alors qu’il commandait en moyenne tous les 15 jours pourrait être marqué comme à risque. Cette méthode repose sur des règles empiriques définies en collaboration avec les équipes marketing et commerciales, qui connaissent les habitudes typiques des clients.

Une fois les clients à risque identifiés, des campagnes ciblées seraient mises en place pour tenter de les réactiver. Cela pourrait inclure des offres promotionnelles, des réductions spécifiques ou des enquêtes de satisfaction pour comprendre les raisons de leur désengagement. Les résultats de ces campagnes seraient ensuite suivis manuellement en comparant les taux de réponse et les comportements observés après les interventions.

Enfin, on procéderait à une évaluation des performances basée sur des indicateurs simples comme la variation du taux de churn avant et après les actions menées, ainsi que sur le chiffre d’affaires récupéré grâce à la rétention des clients identifiés.

Cette approche manuelle, bien que fonctionnelle pour des ensembles de données de petite taille, présente plusieurs limites. Elle est lente, nécessite une surveillance constante et devient rapidement inefficace face à des bases de données volumineuses comme celle utilisée ici. De plus, elle repose sur des hypothèses fixes et ne prend pas en compte les relations complexes ou non linéaires entre les variables.

C’est pourquoi une solution basée sur des modèles prédictifs automatisés est généralement préférée. Elle permet d’identifier plus rapidement les tendances, d’adapter les seuils dynamiquement et d’assurer un suivi en temps réel des comportements clients.

**Listes d’hypothèse**

Hypothèses techniques :

1. Relation entre les variables et le churn : Les comportements d’achat (fréquence, valeur moyenne, délai entre les achats) sont des indicateurs fiables pour prédire le risque de churn.
2. Distribution des données : Les variables disponibles suivent une distribution cohérente et suffisamment représentative pour permettre l’entraînement d’un modèle supervisé.
3. Indépendance des variables : Les caractéristiques analysées, comme la région et les stratégies de rétention appliquées, influencent le churn indépendamment les unes des autres, ou leurs corrélations peuvent être gérées par le modèle.
4. Stabilité des comportements : Les tendances identifiées dans les données historiques se maintiennent dans le futur, rendant les prédictions valides sur les nouvelles observations.
5. Qualité des données : Les informations fournies sont complètes, sans erreurs significatives ni biais qui pourraient fausser les résultats.
6. Performance des modèles : Un modèle supervisé basé sur des algorithmes tels que la régression logistique ou XGBoost sera capable de capturer efficacement les relations complexes entre les variables.
7. Seuils de classification : Une probabilité de churn supérieure à 0,5 est un indicateur suffisant pour déclencher des actions commerciales.

Hypothèses métiers :

1. Signaux de désengagement : Une baisse de fréquence des achats ou un allongement des délais entre les commandes reflètent un risque accru de churn.
2. Impact des stratégies de fidélisation : Les programmes de fidélité et les remises réduisent efficacement la probabilité de churn pour les clients à risque.
3. Valeur client : Les clients avec une valeur vie élevée (Lifetime\_Value) sont plus rentables et doivent être priorisés dans les campagnes de rétention.
4. Comportements régionaux : Les habitudes d’achat peuvent varier selon la région, nécessitant des approches personnalisées pour chaque segment géographique.
5. Influence de la saisonnalité : Les pics de ventes observés autour des saisons spécifiques influencent temporairement les comportements, mais les clients réguliers devraient maintenir un certain niveau d’activité en dehors de ces périodes.
6. Temporalité des actions : Une intervention rapide (avant que l’inactivité devienne chronique) est essentielle pour maximiser les chances de réactivation des clients à risque.

**Possibilité des hypothèses**

J'ai vérifié plusieurs hypothèses en analysant les distributions, les corrélations et les tendances temporelles dans les données.

Les résultats montrent des relations intéressantes entre certaines variables et la probabilité de churn. Par exemple, les graphiques de distribution indiquent des tendances dans la fréquence des achats, la valeur moyenne des commandes et le délai entre les achats. De plus, la relation entre l’ancienneté des clients et leur probabilité de churn semble confirmer que le temps depuis le lancement influence ce risque.

J’ai également mis à disposition un tableau détaillant les corrélations avec le churn, ainsi que des statistiques descriptives pour vérifier les valeurs aberrantes. Nous pourrons approfondir ces résultats et ajuster les hypothèses si nécessaire en fonction des observations spécifiques.

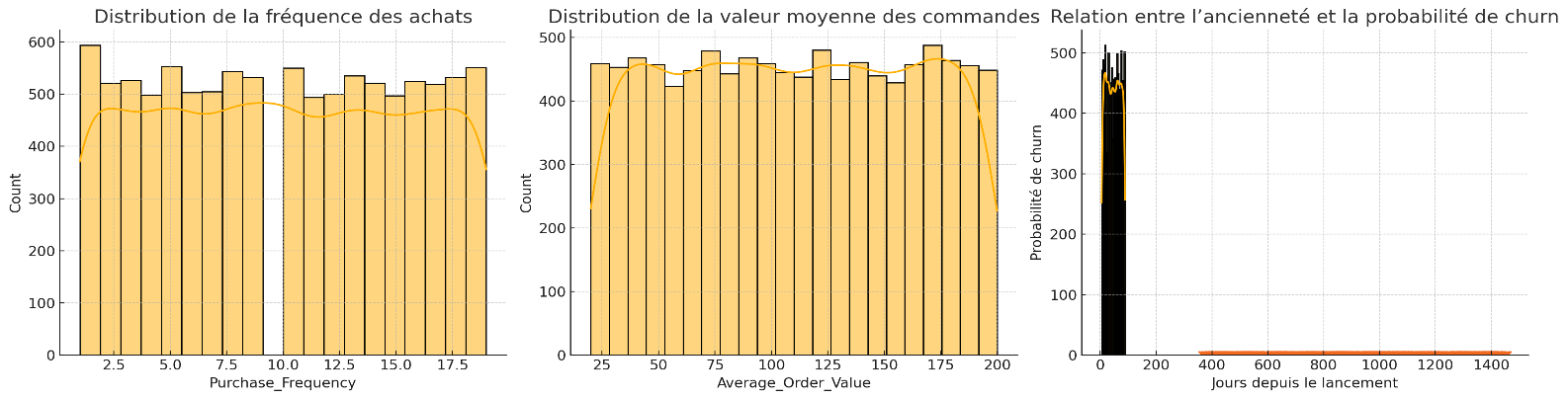


Tableau 1 : Corrélation avec Churn

|  |  |
| --- | --- |
| **Variable** | **Corrélation avec Churn** |
| Churn\_Probability | 1.000000 |
| Purchase\_Frequency | 0.018417 |
| Time\_Between\_Purchases | 0.010410 |
| Days\_Since\_Launch | 0.001371 |
| Average\_Order\_Value | -0.000871 |
| Lifetime\_Value | -0.006847 |

Tableau 2 : Statistiques des Valeurs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistique** | **Fréquence d'achat** | **Valeur Moyenne** | **Temps entre Achats** | **Probabilité de Churn** |
| count | 10000.000000 | 10000.000000 | 10000.000000 | 10000.000000 |
| mean | 9.955700 | 110.006022 | 46.885300 | 0.501552 |
| std | 5.526209 | 52.096870 | 24.711746 | 0.288289 |
| min | 1.000000 | 20.010000 | 5.000000 | 0.000000 |
| 25% | 5.000000 | 65.262500 | 25.000000 | 0.250000 |
| 50% | 10.000000 | 109.930000 | 47.000000 | 0.500000 |
| 75% | 15.000000 | 155.480000 | 68.000000 | 0.750000 |
| max | 19.000000 | 199.960000 | 89.000000 | 1.000000 |