**Commentaires sur les projets contenus dans le portfolio**

**Partie 1** : **Analyse exploratoire des données**

Les objectifs de cette partie sont les suivantes :

* Comprendre la structure des données et leur distribution ;
* Identifier les variables importantes et les relations entre elles ;
* Nettoyer si possible les données en cas d’impuretés ;
* Produire des visualisations claires pour présenter les résultats.

Les notions en lien avec cette partie, seront parcourues en traitant de plusieurs problématiques en lien avec les secteurs de la finance, du marketing, des télécommunications et bien d’autres. Pour chaque étude, nous présentons les aspects suivants :

* Un bref contexte sur la problématique traitée ;
* Les principales tendances en mettant un ancrage sur celles en lien avec la problématique ;
* Les données utilisées pour répondre à cette problématique ;
* Les technologies utilisées dans le cadre de cette étude ;
* Les compétences acquises.

**NB :** Cette partie sera mise en évidence dans les prochains projets de data science au vu de son importance.

**Problématique 1 :** **Comment analyser les transactions bancaires pour identifier les schémas de fraude ?**

Les concepts clés qui apparaissent dans cette problématique sont les suivantes : Transactions bancaires, les schémas de fraude, analyse des données transactionnelles, anomalies, et les technologies de détection de la fraude.

**Définition des concepts clés en lien avec la problématique**

* **Transactions bancaires :** elles représentent l’ensemble des opérations financières effectuées via un système bancaire, telles que : les virements bancaires, paiements par carte, retraits d’espèces, dépôts ou prélèvements automatiques (Mendes & al. (2020), Ferreira (2019)).
* **Schémas de fraude :** D’après l’ Association of Certified Fraud Examiners (ACFE, 2022), ils désignent des modèles ou des comportements répétitifs qui visent à contourner les règles et à réaliser des gains financiers de manière illégale. Parmi les exemples de schémas courants, nous pouvons citer :
* Fraude par carte bancaire : Utilisation non autorisée des cartes pour effectuer des transactions.
* Blanchiment d’argent : Dissimuler l’origine illicite de fonds via des opérations complexes.
* Fraude par phishing : Vol des identifiants bancaires via des moyens frauduleux pour initier des transactions non autorisées.
* Fraude interne : Actes frauduleux perpétrés par des employés de l’institution.
* **Analyse des données transactionnelles :** Selon les études de Nguyen & al. (2021) et celles de Phua & al. (2010), cela consiste à examiner les flux d’informations liées aux transactions afin de repérer des anomalies ou des motifs caractéristiques de fraude. Cela implique :
* Techniques d’analyse exploratoire : Identifier les irrégularités dans les données volumineuses.
* Détection d’anomalies : repérer des transactions qui ne correspondent pas aux comportements habituels.
* **Anomalies :** représentent des écarts significatifs par rapport à des comportements transactionnels normaux (exemple : des virements inhabituels dans un laps de temps très court, ou une transaction effectuée dans un pays éloigné de la résidence habituelle) (Zou & al. (2018), Mendes & al. (2020)). Ces anomalies constituent souvent des indicateurs préliminaires de fraude.
* **Technologies de détection de la fraude**

Selon IBM (2023), Nguyen & al. (2021) et Youssef & al. (2019),elles incluent :

* Systèmes basés sur des règles : ils reposent sur des critères prédéfinis (exemple : blocage automatique des transactions dépassant un certain seuil) ;
* Les systèmes d’intelligence artificielle (IA) : les algorithmes identifient des modèles de fraude en analysant des historiques de transactions et en apprenant à distinguer les comportements normaux des comportements suspects. Par exemple, le clustering (groupement) pour détecter des groupes de transactions inhabituels.
* Systèmes d’intelligence artificielle (IA) : Ces systèmes intègrent des techniques avancées pour détecter des schémas complexes, y compris des fraudes émergentes.

**Contexte :** Les transactions bancaires sont au cœur des activités économiques mondiales, facilitant les échanges financiers entre individus, entreprises et institutions. Toutefois, leur volume et leur complexité croissants offrent également un terrain propice à des activités frauduleuses. La fraude bancaire peut prendre de nombreuses formes, telles que les transactions non autorisées, les détournements de fonds, les activités de blanchiment d'argent ou les escroqueries organisées (Mendes et al., 2020)

Face à ce phénomène, les institutions financières sont confrontées à des défis de taille : comment identifier rapidement les anomalies dans un flux massif de transactions et repérer des schémas de fraude souvent subtils, adaptatifs et en constante évolution ? Selon une étude de Deloitte (2022), les institutions financières perdent chaque année des milliards de dollars à cause de la fraude, et les approches traditionnelles de détection reposant uniquement sur des règles prédéfinies sont souvent dépassées.

Pour répondre à ces défis, l’analyse des données transactionnelles joue un rôle clé. Elle s’appuie sur des techniques avancées, comme le machine learning et l’intelligence artificielle, pour identifier des schémas de fraude, même lorsqu’ils sont dissimulés dans des comportements apparemment normaux. L’’etude de Nguyen et al. (2021) a montré que l’utilisation des algorithmes d’apprentissage supervisé et non supervisé permet de détecter des anomalies dans les transactions en temps réel, tout en s’adaptant aux nouvelles stratégies employées par les fraudeurs.

En complément, les schémas de fraude sont souvent analysés à travers des outils spécifiques tels que les systèmes de détection d’anomalies, les algorithmes de clustering et les réseaux bayésiens, qui permettent de modéliser des relations complexes entre différentes variables transactionnelles. Ces outils permettent non seulement de détecter les fraudes, mais aussi de fournir des alertes précises et d'améliorer les mécanismes de prévention.

Dans ce contexte, il devient crucial pour les banques et les institutions financières d’explorer les technologies émergentes et de concevoir des modèles d’analyse robustes. Ces modèles doivent non seulement être précis, mais aussi capables de traiter de grands volumes de données tout en réduisant les faux positifs, qui sont une source de coûts opérationnels importants (IBM Global Fraud Impact Report, 2023).

Ainsi, l’analyse des transactions bancaires pour détecter les schémas de fraude représente non seulement une priorité stratégique pour les institutions financières, mais également un défi technologique et organisationnel nécessitant des approches multidimensionnelles combinant expertise métier, science des données et intelligence artificielle.

Avant de passer aux principales tendances du secteur de la finance, nous fournissons quelques références qui couvrent les transactions bancaires, la détection des anomalies, les schémas de fraude et l’utilisation des technologies avancées.

**Références bibliographiques :**

Deloitte. (2022). *Global banking fraud survey: Trends and challenges in fraud prevention*. Deloitte Insights.

Cette référence traite des défis et des tendances en matière de détection et prévention des fraudes dans le secteur bancaire. Elle met l’accent sur l'impact financier de la fraude et l’évolution des techniques de prévention.

IBM. (2023). *Global Fraud Impact Report 2023*. IBM Security.

Ce rapport analyse l’impact global de la fraude financière, les outils technologiques comme l’IA, et les défis associés à la gestion des faux positifs et à la détection en temps réel.

Mendes, R., Silva, J., & Ferreira, P. (2020). Fraud detection in banking transactions using machine learning. *Journal of Financial Analytics*, 12(4), 45-59.

Cette publication explore spécifiquement l’utilisation de techniques de machine learning pour détecter des fraudes dans les transactions bancaires.

Nguyen, T. H., Patel, S., & Zhang, Y. (2021). Machine learning approaches for detecting fraudulent transactions: A comparative study. *International Journal of Data Science and Analytics*, 9(3), 125-140.

Cet article compare différentes approches d’apprentissage automatique (supervisé et non supervisé) pour identifier des anomalies dans les données transactionnelles.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*, 34(3), 1-14.

Cet article de revue analyse les techniques de data mining et d’analyse transactionnelle appliquées à la fraude financière.

Youssef, M., & Alzahrani, A. (2019). Big data analytics for detecting financial fraud. *Journal of Information Systems Management*, 35(2), 123-139.

Cette publication définit le rôle des technologies de Big Data dans l’analyse des transactions bancaires et la mise en place de systèmes de détection des fraudes.

Zou, J., & Schiebinger, L. (2018). AI systems in financial fraud detection: Challenges and biases. *Nature Machine Intelligence*, 1(1), 20-27.

Cet article explore l’impact de l’intelligence artificielle sur la détection des fraudes, en détaillant des concepts comme les anomalies transactionnelles et l’optimisation des systèmes de surveillance financière.

Association of Certified Fraud Examiners (ACFE). (2022). *Fraud Prevention Handbook*. ACFE Publications.

Ce guide explique les différents types de fraudes bancaires, les techniques d’identification des schémas de fraude, et les mesures préventives adoptées par les institutions financières.

**Les principales tendances du secteur de la finance en lien avec cette problématique**

Le secteur financier est confronté à des évolutions notables en matière de fraude et de sécurité des paiements. Voici les principales tendances actuelles :

* **Augmentation des fraudes aux transactions bancaires**

Au premier semestre 2024, le nombre d'opérations frauduleuses a augmenté de 12 %, atteignant 3,9 millions de transactions, avec une prédominance des fraudes par carte bancaire représentant 95 % des cas.

<https://www.banque-france.fr/fr/interventions-gouverneur/presentation-des-statistiques-de-fraude-au-1er-semestre-2024-par-losmp>

* **Stabilité du montant total de la fraude**

Malgré l'augmentation du nombre de fraudes, le montant total est resté stable à 1,195 milliard d'euros en 2023, soit une légère hausse de 0,2 %. Cette stabilité est attribuée aux mesures de sécurité renforcées, notamment l'authentification forte introduite par la directive européenne DSP2.

<https://www.lemonde.fr/argent/article/2024/09/11/l-arsenal-contre-la-fraude-au-faux-conseiller-bancaire-bientot-renforce_6312215_1657007.html>

* **Evolution des méthodes de fraude**

Les fraudeurs adoptent des techniques de plus en plus sophistiquées, telles que le « vishing » (hameçonnage vocal) et le « smishing » (hameçonnage par SMS), souvent amplifiées par l’utilisation de l’intelligence artificielle pour le clonage de la voix et les deepfakes.

<https://www.inetco.com/fr/blog/a-look-back-at-payment-fraud-trends-of-2023/>

* **Défis liés aux paiements en temps réel**

La rapidité accrue des paiements en temps réel complique la détection des transactions frauduleuses, réduisant le temps disponible pour les analyses et la surveillance en temps réel.

<https://www.ey.com/fr_ca/insights/payments/lessons-in-managing-real-time-payments-fraud>

* **Renforcement des mesures de sécurité**

Face à la montée des anarques en ligne, les banques intensifient leurs efforts pour sécuriser les transactions, informer les clients des risques et améliorer la détection des activités suspectes.

<https://www.village-justice.com/articles/arnaque-ligne-banques-quelles-sont-les-obligations-des-etablissements-bancaires%2C51816.html>

* **Initiatives légales et réglementaires**

Des mesures législatives sont mises en place pour renforcer la lutte contre la fraude, notamment l’obligation pour les opérateurs de bloquer les appels non authentifiés affichant des numéros usurpés depuis le 1er octobre 2024.

<https://www.lemonde.fr/argent/article/2024/10/24/escroquerie-bancaire-par-spoofing-la-bnp-definitivement-condamnee-a-rembourser-une-victime_6358941_1657007.html>

Ces tendances soulignent la nécessité pour le secteur financier de continuer à innover et à renforcer ses stratégies de prévention et de détection de la fraude afin de protéger les consommateurs et de maintenir la confiance dans les systèmes de paiement.

**Description des données utilisées pour répondre à cette problématique**

**Dataset :** [**https://www.kaggle.com/datasets/mlgulb/creditcardfraud**](https://www.kaggle.com/datasets/mlgulb/creditcardfraud)

**Problématique 2 :** **Quels sont les segments de clients les plus rentables en fonction de leurs interactions avec des campagnes ?**

**Définition des concepts clés en lien avec la problématique**

* **Segmentation des clients :** renvoie au processus de division d’un marché en sous-groupes homogènes de consommateurs partageant des caractéristiques ou des comportements similaires. Elle permet d’adapter les stratégies de marketing aux besoins spécifiques de chaque segment afin d’améliorer l’efficacité des campagnes et la rentabilité. Wedel & al. (2012) définissent la segmentation comme une approche méthodologique visant à identifier des groupes distincts de consommateurs en fonction des critères comportementaux, démographiques ou psychographiques.
* **Rentabilité des segments clients :** fait référence à la valeur économique qu’un groupe spécifique de consommateurs apporte à l’entreprise. Elle est souvent évaluée à travers indicateurs financiers comme la valeur à vie du client (Customer Lifetime Value – CLV) ou la marge brute générée par chaque segment. Gupta & al. (2006) soulignent que la rentabilité des segments repose sur leur capacité à générer des revenus récurrents tout en minimisant les couts d’acquisition et de fidélisation.
* **Interaction avec les campagnes marketing :** englobent l’ensemble des actions entreprises par les clients en réponse aux efforts promotionnels d’une entreprise (clics sur des mails, visites de sites web, engagement sur les réseaux sociaux, achats, etc.). Lemon & al. (2016) expliquent que ces interactions sont essentielles pour comprendre le parcours client et ajuster les stratégies de communication. Elles permettent d’évaluer la pertinence des campagnes et d’adapter les messages pour maximiser l’engagement et la conversion.
* **Analyse RFM (Récence, Fréquence, Montant) :** est une méthode permettant de classer les clients en fonction de trois critères : la récence de leur dernier achat (délai écoulé depuis le dernier achat), la fréquence de leurs achats et le montant dépensé. Fader & al. (2009) affirment que cette approche est particulièrement efficace pour identifier les clients à fort potentiel et orienter les stratégies marketing vers les segments les plus rentables.
* **Customer Lifetime Value (CLV) :** est un indicateur qui mesure la valeur financière qu’un client apportera à une entreprise sur l’ensemble de sa relation commerciale. Selon Gupta & al. (2004), le CLV permet de prédire la rentabilité future des clients et d’ajuster les investissements marketing en conséquence. Il est essentiel pour allouer les ressources aux segments offrant le meilleur retour sur investissement.
* **Machine Learning et segmentation comportementale :** L’utilisation du machine learning pour la segmentation comportementale consiste à appliquer des algorithmes d’apprentissage automatique afin d’identifier des schémas d’interaction et de prédire les comportements futurs des clients. Kumar & al. (2018) expliquent que cette approche permet de personnaliser les offres marketing et d’optimiser les campagnes en fonction des segments les plus réactifs.

La segmentation des clients et l’analyse de leur rentabilité sont des éléments essentiels pour optimiser les campagnes marketing. Grace aux avancées en intelligence artificielle et aux analyses comportementales, les entreprises peuvent affiner leurs stratégies et maximiser leur retour sur investissement.

**Contexte :**

Dans un environnement concurrentiel où les entreprises investissent massivement dans le marketing digital, il devient essentiel de comprendre quels segments de clients génèrent le plus de valeur. L'objectif est d’identifier les clients qui interagissent activement avec les campagnes et qui présentent un fort potentiel de rentabilité. Cette analyse permet d’optimiser l’allocation des ressources marketing et d’améliorer le retour sur investissement des campagnes publicitaires. Selon Kotler & al. (2022), une segmentation efficace permet d’affiner la personnalisation des offres et d’adresser des messages pertinents aux bons profils de consommateurs.

Traditionnellement, les entreprises utilisaient des critères démographiques simples pour segmenter leur clientèle. Cependant, avec l’essor du **Big Data** et de l'**intelligence artificielle**, l'analyse des interactions des clients avec les campagnes (taux d’ouverture des emails, clics sur les publicités, comportements d’achat, etc.) offre une approche plus précise et dynamique. Wedel & al. (2012) soulignent que la segmentation basée sur le comportement permet de mieux comprendre les préférences et les attentes des clients, facilitant ainsi l’adaptation des campagnes.

L’un des principaux défis est d’identifier les **clients à forte valeur** à travers des métriques comme le **Customer Lifetime Value (CLV)**, qui évalue la rentabilité à long terme d’un client. Gupta & al. (2006) mettent en évidence que les entreprises qui segmentent leurs clients selon leur valeur future obtiennent de meilleurs résultats financiers. En parallèle, des méthodes comme l’**analyse RFM (Récence, Fréquence, Montant)** permettent de classer les clients en fonction de leur engagement avec les campagnes marketing (Fader & al., 2009).

L’optimisation des campagnes repose également sur l’exploitation des techniques de **machine learning** pour analyser les schémas comportementaux et prédire les segments les plus réceptifs. Kumar & al. (2018) expliquent que l’utilisation de l’apprentissage automatique permet d’affiner la segmentation et d’anticiper les tendances du marché. Grâce à ces modèles prédictifs, les entreprises peuvent adapter en temps réel leurs offres et maximiser l’impact de leurs actions marketing.

En conclusion, l’identification des segments clients les plus rentables est un enjeu stratégique pour les entreprises cherchant à **optimiser leurs campagnes marketing**. L’intégration de technologies avancées et l’analyse des comportements clients permettent d’améliorer la personnalisation des offres et d’augmenter la fidélisation. En s’appuyant sur des référentiels académiques et des études de marché, les entreprises peuvent ainsi mieux comprendre leurs consommateurs et prendre des décisions marketing basées sur des données tangibles.

Les références académiques citées offrent un cadre solide pour approfondir ces concepts et leur application dans le marketing moderne.

**Bibliographie :**

Fader, P. S., & Hardie, B. G. (2009). *Customer-Base Analysis: Dissecting RFM for Customer Value Insights*. *Journal of Marketing Research, 46*(2), 218-232.

Cet article explore l’analyse de la base client à travers la méthode RFM (Récence, Fréquence, Montant) pour mieux comprendre la valeur des clients et affiner les stratégies marketing.

Gupta, S., & Zeithaml, V. (2006). *Customer Metrics and Their Impact on Financial Performance*. *Marketing Science, 25*(6), 718-739.

Les auteurs examinent comment les métriques clients (CLV, acquisition, rétention) influencent la performance financière des entreprises et leur rentabilité à long terme.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). *Valuing Customers*. *Journal of Marketing Research, 41*(1), 7-18.

Cet article propose une approche quantitative pour évaluer la valeur d’un client et son impact sur les décisions marketing et stratégiques des entreprises.

Kotler, P., Keller, K. L., & Chernev, A. (2022). *Marketing Management* (16e éd.). Pearson.

Cet ouvrage de référence en marketing management couvre les concepts clés du marketing stratégique et opérationnel, en mettant l’accent sur l'innovation et l’adaptation aux nouvelles tendances.

Kumar, V., & Reinartz, W. (2018). *Customer Relationship Management: Concept, Strategy, and Tools* (3e éd.). Springer.

Ce livre examine la gestion de la relation client (CRM) en combinant théorie et applications pratiques pour optimiser la fidélisation et la rentabilité des clients.

Lemon, K. N., & Verhoef, P. C. (2016). *Understanding Customer Experience Throughout the Customer Journey*. *Journal of Marketing, 80*(6), 69-96.

L’article analyse le concept d’expérience client tout au long du parcours d’achat, en mettant en avant les interactions clés influençant la satisfaction et la fidélité.

Wedel, M., & Kamakura, W. A. (2012). *Market Segmentation: Conceptual and Methodological Foundations* (2e éd.). Kluwer Academic Publishers.

Les auteurs présentent une approche détaillée de la segmentation du marché, en intégrant des méthodes analytiques et statistiques pour une segmentation efficace et stratégique.

**Les principales tendances du secteur du marketing en lien avec la problématique**

A la lumière des évolutions récentes, plusieurs tendances majeures se dessinent dans le secteur du marketing en 2025 :

* Intelligence Artificielle (IA) et personnalisation avancée : L’IA joue un rôle central en permettant une personnalisation accrue des campagnes marketing. Elle facilite l’analyse en temps réel de vastes ensembles de données, offrant ainsi des expériences client sur mesure. <https://elpais.com/economia/formacion/2025-01-24/la-ia-la-personalizacion-y-la-interaccion-en-redes-sociales-lideran-las-tendencias-de-marketing-digital-en-2025.html>
* Marketing durable et économie circulaire : Les consommateurs privilégient de plus en plus les marques engagées dans des pratiques écoresponsables. L’intégration de l’économie circulaire dans les stratégies de marketing, par le biais de la réduction des déchets et du recyclage, est devenue essentielle.

<https://www.marketing-management.io/blog/tendances-marketing-2025>

* Commerce conservationnel et chatbots : L’utilisation de chatbots alimentés par l’IA transforme les interactions client, offrant des réponses instantanées et améliorant l’expérience utilisateur. Cette tendance renforce l’importance du commerce conversationnel dans les stratégies marketing.

<https://www.marketing-management.io/blog/tendances-marketing-2025>

* Authenticité et transparence des marques : Les consommateurs recherchent des marques authentiques et transparentes, incarnant des valeurs fortes. Le marketing devient plus éthique, centré sur des actions concrètes, renforçant ainsi la confiance et la fidélité des clients.

<https://www.codeur.com/blog/tendances-marketing/>

* Réalité Augmentée (RA) et Réalité Virtuelle (RV) : L’accessibilité croissante de la RA et de la RV permet aux marques de créer des expériences immersives et interactives, captant ainsi l’audience de manière innovante.

<https://www.codeur.com/blog/tendances-marketing/>

* Marketing d’influence et collaboration stratégiques : Les partenariats avec des influenceurs et des personnalités publiques restent une stratégie clé. Les collaborations audacieuses et inattendues génèrent un fort engagement et élargissent la portée des campagnes marketing.

<https://www.blogdumoderateur.com/marketing-digital-tendances-bonnes-pratiques-adopter-2025/>

* Contenu généré par les employés et communautés de marque : Encourager les employés à créer du contenu et á devenir des ambassadeurs de la marque renforce l’authenticité et l’engagement. De plus, le marketing communautaire, qui valorise les contributions des clients fidèles, gagne en importance.

<https://www.cbnews.fr/etudes/image-vistaprint-9-tendances-2025-89604>

Ces tendances illustrent l'évolution du marketing vers une approche plus technologique, éthique et centrée sur l'humain, répondant aux attentes croissantes des consommateurs en matière de personnalisation, de responsabilité sociale et d'innovation.

**Description des données utilisées pour répondre à cette problématique**

**Dataset :** [**https://www.kaggle.com/datasets/rodsaldanha/arketingcampaign/data**](https://www.kaggle.com/datasets/rodsaldanha/arketingcampaign/data)