

Intégration de données dans un datawarehouse

RAPPORT SAE

03/03/2025



Abdelkader Alahoum
Assia Boudjraf
Anisse Elayeb
Ahmed Oubourhim
BUTSD2 VCOD AFA



TABLE DES MATIERES

I.PRÉSENTATION DU PROJET	2
II.CHOIX DU PROCESSUS MÉTIER À MODÉLISER	4
III.GRANULARITE ET DIMENSIONS POUR LE MODELE DE DONNEES	5
IV.IDENTIFICATION DES FAITS	8
V.INTEGRATION DES DONNEES	10
VI.OBSERVATIONS	12
VII.CONCLUSION	14

I - PRÉSENTATION DU PROJET

Le projet porte sur la gestion des données d'achats et de ventes d'une boutique localisé en France (électroménager, bureautique entre autres). On structurera les données pour faciliter l'analyse et la prise de décision. Le modèle adopté permet d'examiner les performances commerciales, les tendances d'achat et d'optimiser les stratégies marketing.

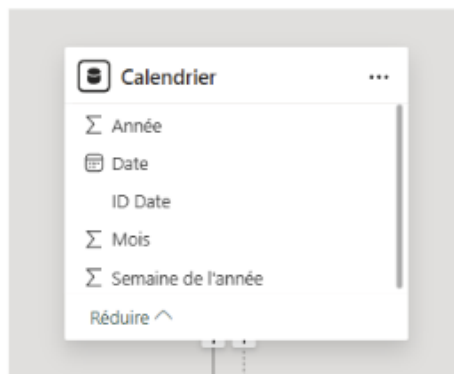
L'objectif principal est de fournir une architecture efficace qui permettra d'extraire des informations pertinentes à partir des données disponibles afin d'aider à la prise de décision stratégique. Cette démarche implique la collecte, le traitement et l'analyse des transactions clients pour optimiser les performances de vente et identifier les tendances de consommation.

Pour garantir une analyse complète et précise, plusieurs sources de données ont été intégrées dans ce modèle :

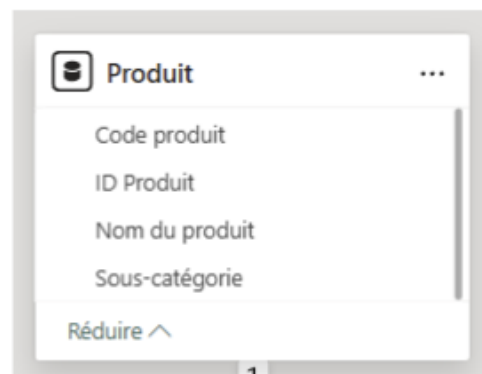
- **Données des transactions d'achat et de vente:** Comprend les informations sur les commandes effectuées, les montants, les dates d'achat.
- **Données clients:** Informations sur les acheteurs incluant leur identification, leur emplacement géographique et leur segment de marché.
- **Données produits :** Caractéristiques des articles vendus, comprenant les catégories de produits, les sous-catégories et les références produits.
- **Données temporelles :** Un calendrier détaillé permettant d'effectuer des analyses à différents niveaux de granularité (jours, mois, années).
- **Données géographiques :** Emplacement des transactions et analyse des tendances d'achat selon les zones géographiques.

SCHEMA

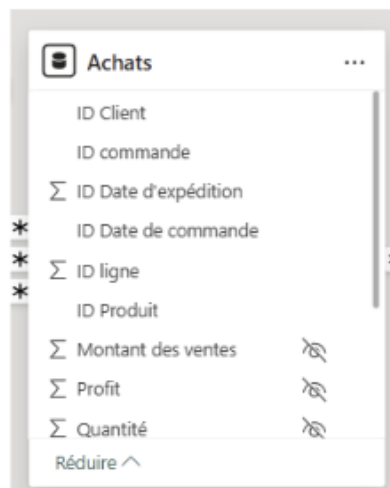
Données temporelles :



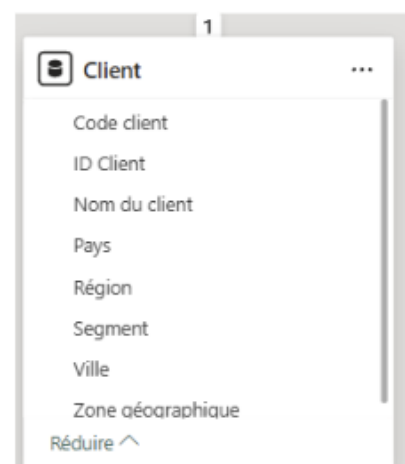
Données produits :



Données de
transactions
d'achat ou vente :



Données clients :



PROBLÉMATIQUE

Comment structurer, modéliser et analyser efficacement les données d'achats et de ventes afin d'améliorer la compréhension des comportements d'achat des clients, d'optimiser la gestion des stocks et de maximiser les performances commerciales dans un environnement concurrentiel ?

II - CHOIX DU PROCESSUS MÉTIER À MODÉLISER

a) Le processus métier choisi

Le processus métier retenu pour ce projet est la gestion des ventes de matériel bureautique. Il s'agit d'analyser les tendances de consommation afin d'identifier les produits les plus populaires, les clients les plus fidèles et les régions où les ventes sont les plus performantes. Cette approche permet d'optimiser les stocks en fonction de la demande réelle et d'adapter les stratégies marketing pour maximiser le chiffre d'affaires.

L'analyse du processus de vente inclut plusieurs aspects clés :

- Gestion des commandes et des transactions : suivi des volumes de ventes et des produits commandés.
- Identification des tendances de consommation : détection des variations saisonnières et des évolutions du marché.
- Segmentation des clients : classification des clients en fonction de leur fréquence d'achat et de leur panier moyen.
- Optimisation des stocks : ajustement des niveaux de stock en fonction de la demande et des prévisions de ventes.

Cette analyse permettra d'améliorer la gestion des ressources, de proposer des offres plus attractives et d'assurer une meilleure satisfaction client.

b) Justification du choix du processus métier

Dans une boutique spécialisée en matériel bureautique, il est essentiel de comprendre les dynamiques de vente afin d'optimiser la rentabilité et la satisfaction client. L'analyse des transactions permet de répondre à plusieurs enjeux stratégiques :

1. Maximisation du chiffre d'affaires
Identifier les produits les plus rentables permet d'adapter l'offre et de concentrer les efforts sur les articles les plus demandés.
2. Optimisation des stocks et de la logistique
Une meilleure connaissance des tendances d'achat évite les ruptures de stock et réduit les coûts liés aux surplus de marchandises.
3. Personnalisation des stratégies marketing
En analysant les comportements d'achat, il devient possible de cibler les promotions et les campagnes publicitaires sur les clients les plus pertinents.
4. Adaptation aux variations saisonnières
Certaines périodes de l'année sont plus propices aux ventes (rentrée scolaire, fin d'année). L'étude des tendances permet d'anticiper ces pics d'activité et d'adapter les offres en conséquence.
5. Amélioration de la fidélisation client
En identifiant les clients les plus fidèles et ceux à fort potentiel, des programmes de fidélisation peuvent être mis en place pour renforcer leur engagement.

L'intégration de ces données dans un SID permet de structurer l'ensemble des informations liées aux ventes et d'obtenir des indicateurs pertinents pour la prise de décision.

III- GRANULARITE ET DIMENSIONS POUR LE MODELE DE DONNEES

D'après le schéma, la base de données concerne la gestion des achats et ventes avec des entités comme les produits, les clients, et les dates. En suivant l'exemple du second document, voici une proposition des dimensions et des mesures associées :

1. Dimensions identifiées:

- **Temps** : Année, Mois, Jour des commandes et expéditions.
- **Client**: ID Client, Nom, Pays, Région, Ville, Segment.
- **Produit**: ID Produit, Nom du Produit, Code Produit, Sous-catégorie.
- **Achats/Ventes** : ID Commande, ID Ligne, ID Produit, ID Client.

2. Tableau croisé des processus (mesures) avec leurs dimensions potentielles:

DIMENSION	PROCESSUS
GESTION DES ACHATS ET /VENTES	
CLIENT	(IDclient,Nom,Pay,Région,ville)
TEMPS	(Date ,mois,jour,année)
PRODUIT	(Produit ID,Nom,Code_produit)
ACHAT	(ID_commande,Montant des ventes ,Quantité)

3. Granularité du modèle:

Pour assurer une analyse précise des données, il est nécessaire de définir un niveau de granularité adapté. Chaque transaction enregistrée dans la base de données représente une ligne de commande unique, ce qui signifie que l'analyse peut être effectuée à un niveau très détaillé. La granularité temporelle est définie au jour près, permettant d'agréger les données à des niveaux supérieurs tels que le mois ou l'année. De même, les informations sur les clients et les produits sont enregistrées de manière unique, ce qui garantit une segmentation efficace et une analyse approfondie des comportements d'achat.

- **Fait principal:** Chaque enregistrement dans la table **Achats** représente une **ligne de commande unique**.
- **Granularité temporelle:** Jour (ID Date), avec possibilité d'agrégation sur Mois et Année.

- **Granularité des clients:** Un enregistrement par **client unique** avec ses informations géographiques.
- **Granularité des produits:** Un enregistrement par **produit unique**

Précision des Dimensions :

1. Table "Achats" (Table des faits)

Cette table représente les transactions effectuées par les clients. Elle contient des mesures qui permettent d'analyser les ventes.

Colonnes:

- **ID commande** → Identifiant unique de la commande.
- **ID Date d'expédition** → Date à laquelle la commande a été expédiée.
- **ID Date de commande** → Date à laquelle la commande a été passée.
- **ID ligne** → Identifiant unique d'une ligne d'achat (une commande peut contenir plusieurs lignes).
- **ID Produit** → Référence du produit acheté (lié à la table "Produit").
- **Montant des ventes** → Valeur monétaire de la vente.
- **Profit** → Bénéfice réalisé sur cette vente.
- **Quantité** → Nombre d'unités achetées pour ce produit.
- **Remise** → Réduction accordée sur cette ligne d'achat.

2. Table "Client" (Dimension)

Cette table contient des informations sur les clients ayant effectué des achats.

Colonnes :

- **Code client** → Référence unique du client.
- **ID Client** → Identifiant unique du client.
- **Nom du client** → Nom complet du client.
- **Pays** → Pays d'origine du client.
- **Région** → Région où se situe le client.
- **Segment** → Catégorie du client (ex. particulier, entreprise, administration).
- **Ville** → Ville où réside le client.

- **Zone géographique** → Grande zone géographique (ex. Amérique du Nord, Europe, Asie).

3. Table "Produit" (Dimension)

Cette table contient les informations sur les articles vendus par la boutique.

Colonnes :

- **Code produit** → Référence unique du produit.
- **ID Produit** → Identifiant unique du produit.
- **Nom du produit** → Libellé du produit (ex. "Classeur A4", "Stylo Bic Bleu").
- **Sous-catégorie** → Catégorie du produit (ex. Papeterie, Mobilier, Informatique).

4. Table "Calendrier" (Dimension)

Cette table permet d'analyser les ventes selon la période.

Colonnes :

- **Année** → Année de la transaction.
- **Mois** → Mois de la transaction.
- **Semaine de l'année** → Numéro de la semaine dans l'année.
- **Date** → Date exacte de la transaction.
- **ID Date** → Identifiant unique de la date pour relier avec la table des achats.

ETL

1. Extraction des données

L'extraction des données est une étape clé qui consiste à récupérer toutes les informations nécessaires depuis différentes sources. Dans ce projet, les données proviennent de bases transactionnelles, de fichiers CSV et Excel contenant l'historique des ventes. Pour éviter toute perte d'information. L'objectif est de s'assurer que l'entreprise dispose toujours de données récentes et fiables pour guider ses décisions commerciales.

2. Transformation des données

Une fois les données extraites, elles doivent être mises en forme pour être facilement exploitables. Cela commence par un nettoyage en supprimant les informations inutiles et en harmonisant les formats (dates,vente,commande). Certaines colonnes sont restructurées pour faciliter l'analyse : par exemple, un ID de date, afin de mieux comprendre la répartition des commandes. Enfin, des indicateurs sont créés, comme le chiffre d'affaires par produit ou le taux de fidélité des clients, rendant l'analyse plus pertinente et exploitable.

3. Gestion des erreurs et des données manquantes

Travailler avec des données brutes implique souvent des erreurs et des valeurs manquantes. Pour garantir des analyses fiables, un travail de vérification est effectué à chaque étape. Les doublons sont supprimés pour éviter de fausser les résultats, tandis que les valeurs manquantes sont soit remplacées par une estimation soit supprimées si elles ne sont pas exploitables. Un contrôle de cohérence est également mis en place pour détecter d'éventuelles anomalies, comme des prix négatifs ou des commandes sans date. L'objectif est d'obtenir une base de données propre et précise, permettant une interprétation juste et pertinente des résultats.

Création de colonnes personnalisées et conditionnelles

Pour affiner l'analyse des données et faciliter l'interprétation des résultats, plusieurs colonnes personnalisées ont été ajoutées dans notre modèle de données. Par exemple, dans la **table Calendrier**, nous avons créé une colonne indiquant la semaine de l'année à partir des dates des transactions. Cela permet d'analyser plus facilement l'évolution des ventes sur des périodes spécifiques et d'identifier des tendances hebdomadaires, comme les pics de commandes en début ou en fin de mois.

Nous avons également utilisé des calculs pour générer des indicateurs clés. Par exemple, dans Power Query, des **formules de somme** ont été appliquées pour calculer le **montant total des ventes** par client, par produit et par période. Ces colonnes permettent de créer des visualisations claires et de comparer rapidement les performances entre différents segments. Grâce à ces transformations, l'entreprise peut mieux comprendre son activité et ajuster ses stratégies commerciales en fonction des résultats obtenus.

= Table.AddColumn("#Semaine de l'année insérée", "ID Date", each Text.Combine({Date.ToText([Date], "yyyy"), "0", Date.ToText([Date], "%M")},							
Date	1 ² 3 Année	1 ² 3 Mois	1 ² 3 Semaine de l'année	1 ² 3 ID Date			
● Valide 100 %	● Valide 100 %	● Valide 100 %	● Valide 100 %	● Valide 100 %	● Valide 100 %	● Valide 100 %	● Valide 100 %
● Erreur 0 %	● Erreur 0 %	● Erreur 0 %	● Erreur 0 %	● Erreur 0 %	● Erreur 0 %	● Erreur 0 %	● Erreur 0 %
● Vide 0 %	● Vide 0 %	● Vide 0 %	● Vide 0 %	● Vide 0 %	● Vide 0 %	● Vide 0 %	● Vide 0 %

IV - IDENTIFICATION DES FAITS

Le choix du modèle en étoile a été motivé par plusieurs raisons essentielles adaptées au contexte de gestion des achats et ventes. Tout d'abord, ce modèle repose sur une table centrale des faits (Achats) connectée à plusieurs tables de dimensions (Temps, Client, Produit, etc.), ce qui simplifie la structuration des données et facilite leur exploitation analytique. L'un des principaux avantages de cette approche est sa simplicité et efficacité. En limitant le nombre de jointures entre les tables, elle permet d'exécuter des requêtes plus

rapidement, ce qui est essentiel pour des analyses en temps réel et des tableaux de bord interactifs.

De plus, le modèle en étoile est optimisé sur Power BI, qui fonctionnent plus efficacement avec ce type de structure, la clarté du modèle permet aux utilisateurs métiers d'explorer facilement les données. Un autre avantage clé est la facilité d'extension. Si de nouvelles dimensions doivent être ajoutées, cela peut être réalisé sans restructurer entièrement l'entrepôt de données.

a) Faits pertinents :

1. **Nombre total de commandes** : Nombre de commandes passées par période.
2. **Montant total des ventes** : Chiffre d'affaires généré par les ventes de produits bureautiques.
3. **Quantité de produits vendus** : Nombre total d'articles achetés.
4. **Remise moyenne appliquée** : Réduction moyenne accordée par commande ou par produit.
5. **Profit généré** : Différence entre les revenus des ventes et le coût des produits vendus.
6. **Pays ou région avec le plus d'achats** : Analyse géographique des ventes pour identifier les marchés les plus porteurs.
7. **Client le plus fidèle** : Identification des clients ayant effectué le plus de commandes ou ayant le panier moyen le plus élevé.
8. **Produit le plus vendu** : Produits les plus populaires en fonction du volume de ventes.

b) Rôle des faits :

Ces indicateurs permettent de structurer l'entrepôt de données et d'offrir une vision claire des performances de vente. Grâce à ce modèle, il devient possible d'analyser les tendances, d'optimiser les stratégies marketing, de mieux gérer les stocks, et d'adapter les offres commerciales en fonction des besoins réels des clients.

L'entrepôt de données servira ainsi de base solide pour la prise de décision, facilitant la segmentation des clients, l'identification des meilleures opportunités commerciales et l'amélioration de la rentabilité globale de la boutique.

123 ID Date	Date	123 Année	123 Mois	123 Semaine de l'année
<ul style="list-style-type: none"> ● Valide 100 % ● Erreur 0 % ● Vide 0 % 	<ul style="list-style-type: none"> ● Valide 100 % ● Erreur 0 % ● Vide 0 % 	<ul style="list-style-type: none"> ● Valide 100 % ● Erreur 0 % ● Vide 0 % 	<ul style="list-style-type: none"> ● Valide 100 % ● Erreur 0 % ● Vide 0 % 	<ul style="list-style-type: none"> ● Valide 100 % ● Erreur 0 % ● Vide 0 %
1	20200101	01/01/2020	2020	1
2	20200102	02/01/2020	2020	1
3	20200103	03/01/2020	2020	1

L'onglet Affichage nous a permis de contrôler la qualité des données et de tenter d'éliminer autant que possible les irrégularités, telles que le fait que certaines variables numériques étaient traitées comme des caractères. Nous avons donc dû convertir les points en virgules pour que Power BI puisse comprendre nos instructions et ainsi transformer la variable caractère en variable numérique.

V - INTÉGRATION DES DONNÉES (PROCESSUS ETL)

1. Extraction des données

Les données utilisées dans ce projet proviennent principalement de sources disponibles sur GitHub, où elles ont été collectées et structurées pour être intégrées dans notre système d'analyse. Ce choix garantit un accès fiable et centralisé aux informations essentielles concernant les transactions d'achat et de vente, les clients, les produits et les dimensions temporelles et géographiques. L'utilisation d'une plateforme comme GitHub permet également de suivre l'évolution des données et de collaborer efficacement sur leur transformation et leur exploitation.

2. Transformation des données

Avant d'être exploitées, les données brutes doivent être transformées afin d'assurer leur cohérence et leur pertinence. Ce processus comprend plusieurs étapes : suppression des colonnes non pertinentes, conversion des types de données pour garantir leur compatibilité

avec les outils d'analyse, fusion et séparation des colonnes en fonction des besoins analytiques, et normalisation des formats. L'objectif est de structurer les données de manière à faciliter leur exploitation tout en garantissant leur intégrité et leur fiabilité.

3. Gestion des erreurs et des données manquantes

Un contrôle rigoureux est effectué pour identifier et corriger les erreurs dans les données. Cela inclut la suppression des doublons, la gestion des valeurs nulles et la correction des incohérences. Ces traitements sont essentiels pour garantir la qualité des analyses et éviter les biais dans l'interprétation des résultats.

3. Création de colonnes personnalisées

Certaines colonnes personnalisées ont été ajoutées afin d'enrichir l'analyse des données et de faciliter l'interprétation des résultats. L'une des principales colonnes créées est la colonne Calendrier, qui permet d'établir une correspondance précise entre les transactions et leur période d'occurrence. Cette colonne a été générée afin de structurer les analyses temporelles en offrant une vue détaillée des ventes selon les jours, les mois et les années. Grâce à cette colonne, il est possible d'agréger les données en fonction des périodes spécifiques et d'analyser les tendances saisonnières des ventes.

Cette amélioration permet non seulement de mieux comprendre l'évolution des ventes sur une période donnée, mais aussi d'identifier les pics d'activité, ce qui aide à anticiper les besoins en stocks et à ajuster les stratégies marketing en fonction des périodes les plus propices à la consommation.

En complément, d'autres colonnes ont été générées par des commandes DAX pour calculer des indicateurs spécifiques comme le montant total des ventes par transaction, la marge bénéficiaire et l'impact des promotions appliquées sur les ventes. Ces colonnes permettent d'obtenir des insights plus approfondis et de rendre les analyses plus pertinentes. Certaines colonnes personnalisées sont ajoutées pour enrichir l'analyse et faciliter la compréhension des données. Par exemple, des indicateurs calculés permettent d'évaluer la rentabilité des ventes en fonction des marges réalisées, ou encore de mesurer l'impact des promotions sur le comportement des clients.

VI – OBSERVATIONS

Nous avons choisi d'analyser certains aspects clés des ventes afin d'identifier les opportunités d'optimisation et de prise de décision stratégique. Parmi les différentes dimensions possibles, nous avons décidé de nous concentrer sur trois axes majeurs.

1. Répartition géographique des ventes

L'analyse des ventes révèle que la boutique opère exclusivement en Europe, avec une concentration significative dans trois pays principaux :

- **France** : 86 000 unités vendues
- **Royaume-Uni** : 80 000 unités vendues
- **Allemagne** : 78 000 unités vendues

Derrière ces marchés principaux, d'autres pays comme l'Espagne (40K €), l'Italie (36K €) et l'Autriche (18K €) affichent des résultats plus modestes mais restent des opportunités de croissance. Les pays comme les Pays-Bas, la Suisse, la Finlande et la Suède, avec des profits inférieurs à 10K €, représentent des marchés moins significatifs, mais qui pourraient être développés à travers des stratégies adaptées.

Les marchés principaux représentent les sources de revenus les plus considérables pour l'entreprise. Leur importance justifie une attention particulière en termes de stratégies marketing et logistiques, afin de maximiser la performance commerciale sur ces zones géographiques.

2. Analyse des clients

L'étude des clients révèle un groupe de grands acheteurs, dont la fidélité et le volume de commandes jouent un rôle clé dans le chiffre d'affaires de l'entreprise. Les trois clients les plus importants sont :

1. Ignace Thérien avec 149 unités achetées
2. Astrid Gour avec 145 unités achetées
3. Mathilde Gabriaux avec 145 unités achetées

Le top 15 des clients réalise un volume de commandes significatif, avec une consommation comprise entre 111 et 145 unités. L'étude des clients montre également qu'environ 20 clients ont passé plus de 100 commandes. Cette tendance met en avant la nécessité d'une stratégie de fidélisation renforcée, incluant des offres personnalisées, des réductions pour gros volumes ou des services exclusifs, afin de maximiser leur engagement et encourager des achats encore plus fréquents.

3. Analyse des produits les plus vendus

Concernant les produits, on observe que certains articles se démarquent nettement par leur popularité :

1. **Acco couverture de classeur (recyclable) : 30 commandes**
2. **Acco couverture de classeur (économique) : 17 commandes**
3. **Acco perforatrice économie : 15 commandes**

Ces tendances de consommation montrent un attrait marqué pour les fournitures de bureau essentielles, en particulier celles offrant un bon équilibre entre économie et écologie. Afin d'exploiter au mieux ces informations, plusieurs axes stratégiques peuvent être envisagés :

- Anticiper la demande et ajuster les stocks pour éviter les ruptures de ces articles phares, garantissant ainsi une disponibilité constante pour les clients.
- Développer des offres promotionnelles ciblées, notamment sur les produits les plus populaires, afin d'encourager leur achat en volume et de dynamiser encore davantage les ventes.

- Explorer de nouvelles gammes de produits complémentaires, en capitalisant sur l'intérêt des consommateurs pour les fournitures recyclées et économiques, et ainsi élargir l'offre tout en répondant aux attentes du marché.

Cela nous permet de mieux orienter les décisions stratégiques en matière de stocks, de gammes et de promotions pour maximiser les performances commerciales.

VI - CONCLUSION

L'analyse des données de vente et d'achat a mis en évidence plusieurs tendances clés qui constituent des leviers stratégiques pour l'entreprise. Nous avons désormais une vision précise des marchés les plus porteurs, des clients les plus fidèles et des produits les plus prisés. Ces informations sont essentielles pour optimiser la gestion des stocks, renforcer la relation client et affiner les stratégies marketing afin de maximiser les performances commerciales.

Solutions et recommandations

Pour répondre efficacement à la demande et éviter les ruptures de stock, l'entreprise doit ajuster sa gestion des approvisionnements en tenant compte des articles les plus vendus. En effet, les produits les plus populaires, tels que l'Acco couverture de classeur recyclable (30 commandes), l'Acco couverture de classeur économique (17 commandes) et l'Acco perforatrice économique (15 commandes), nécessitent un suivi rigoureux afin d'assurer leur disponibilité continue.

Par ailleurs, la fidélisation de la clientèle représente un axe stratégique majeur. L'analyse des comportements d'achat révèle qu'environ 20 clients ont passé plus de 100 commandes, témoignant d'un engagement fort envers la marque. Pour valoriser cette fidélité et encourager une augmentation des volumes d'achat, la mise en place d'un programme de fidélité apparaît comme une solution pertinente. Une carte offrant des réductions exclusives, des avantages

personnalisés ou des offres promotionnelles sur les produits préférés pourrait renforcer leur attachement à l'entreprise et stimuler leurs achats.

En parallèle, pour accroître sa présence sur le marché international, l'entreprise doit capitaliser sur ses trois principaux marchés : la France (86K unités vendues), le Royaume-Uni (80K) et l'Allemagne (78K). Cependant, une expansion plus large en Europe constitue une opportunité de croissance à ne pas négliger. Pour cela, une stratégie de communication digitale ciblée serait un levier efficace. En intensifiant sa présence sur les réseaux sociaux des pays visés et en adaptant localement ses campagnes publicitaires grâce à des outils comme les VPN, l'entreprise pourrait toucher un public plus large et renforcer son attractivité.

En mettant en œuvre ces actions stratégiques, l'entreprise pourra optimiser sa gestion des ressources, fidéliser ses clients à forte valeur et accélérer son développement international, consolidant ainsi sa position sur le marché.