

[Rapport d'Analyse]

Date : 08/2024

Haïdopoulo Matthieu

**Introduction**

* + **Objectif du rapport**

L'objectif de ce rapport est d'analyser, d'identifier et de comparer différentes méthodes de connexion et de traitement des données, afin de recommander la solution la plus pertinente pour répondre aux besoins de l'entreprise (tels que le suivi des KPI, la gestion des marges, et l'analyse de la performance des nouveaux marchés…), dans le cadre de la création et utilisation d'un outil de data visualisation. Le tableau de bord qui en résultera sera essentiel pour le PDG ainsi que les chefs de produits, leur permettant de piloter l'activité de l'entreprise avec des données fiables et accessibles.

* + **Contexte de l'analyse**

L'entreprise fait face à des défis importants liés à l'utilisation des données. Malgré la récente mise à jour de leurs bases de données, ils sont encore en retard par rapport à leurs concurrents en termes de gestion et d'exploitation des données. Ce retard impacte directement notre compétitivité.

Par ailleurs, ils font face à des difficultés croissantes dans la gestion des stocks, avec des erreurs récurrentes et des problèmes de réapprovisionnement.

Enfin, ils sont dans l’incapacité à identifier clairement les relais de croissance sur les nouveaux segments de marché, notamment en ce qui concerne le choix des produits, la fixation des prix, et la gestion des marges. Dans ce contexte, la mise en place d'une solution de data visualisation devient cruciale pour remédier à ces enjeux et renforcer notre position sur le marché.

**État des lieux et évaluation de la pertinence**

* + **2.1 Description de la situation actuelle**

Suite au travail de nettoyage des données et à la présentation faite au CODIR pour identifier les erreurs dans nos différentes bases de données, le PDG a lancé un projet de mise à jour des bases de données. Les données de l'entreprise sont désormais stockées dans une base de données au format SQLite.

Un schéma relationnel a été fourni pour illustrer les relations entre les différentes tables et les entités qu'elles représentent. Le dictionnaire des données accompagne ces informations, en détaillant les champs et leur signification afin d'assurer une utilisation cohérente et précise des données dans les analyses futures.

Ils se présentent comme suit :

* **Tables** :
  + **web** : Catalogue de produits. Contient les descriptions des produits (nom, référence SKU, contenant, type de produit, etc.) ainsi que les informations liées au site web (date de création et de modification, URL du produit, etc.).
  + **finance** : Informations sur les produits, incluant les prix d'achat et de vente par mois, et si les produits sont vendus sur le site web ou non.
  + **promo** : Détails des promotions des produits, y compris les dates de début et de fin, le prix de vente en promotion, et le nombre de ventes réalisées pendant la promotion.
  + **sales** : Données sur les ventes hors promotion, les stocks au dernier jour du mois, et le nombre de produits commandés chez les fournisseurs au cours du mois.
* **Relations** :
  + Les tables **finance**, **promo**, et **sales** sont reliées via la clé Key\_finance, une clé primaire unique constituée par la concaténation de la date (année-mois) et de la référence produit.
  + La table **web** est reliée à la table **finance** via la clé SKU, correspondant à la référence des produits.
  + **2.2 Évaluation de la pertinence des données existantes**

Pour répondre au besoin de création d'un tableau de bord, les données actuelles couvrent de manière adéquate les aspects commerciaux et l'analyse des ventes ou des stocks. Elles permettent de suivre les performances des produits, les marges, et l'efficacité des promotions. Cependant, pour enrichir l'analyse par une segmentation fine de la clientèle et un ciblage précis, des informations supplémentaires seraient nécessaires, notamment des données plus détaillées sur les clients.

**Qualité et complétude des données**

L'analyse des données a mis en lumière plusieurs points concernant leur qualité et leur adéquation aux besoins de visualisation :

* **Types de données et formats** :
  + Une erreur de concordance a été identifiée pour la variable SKU, qui est au format **INT** dans la table web et au format **VARCHAR** dans la table finance.
  + La variable sales\_offer de la table promo, actuellement au format **INT**, devrait être convertie au format **FLOAT** pour refléter plus précisément les prix.
  + Certaines variables pourraient être converties en type **booléen**, bien que cela ne soit pas impératif.
* **Doublons** :
  + Les clés primaires sont uniques, ce qui garantit l'intégrité des identifiants. Cependant, deux doublons ont été trouvés dans la variable post\_title (SKU 14000 – 14828 et 15669 – 16034). Le premier est un doublon avéré, tandis que le second pourrait correspondre à deux produits différents, mais cette distinction n'est pas claire dans les données.
* **Valeurs manquantes et nulles** :
  + Il n'y a aucune valeur manquante ou nulle dans les données.
* **Valeurs distinctes et cohérence** :
  + Aucune incohérence orthographique n'a été détectée dans les valeurs.
  + Il n'y a aucune incohérence entre les variables product\_type et segment, chaque product\_type étant correctement associé à un segment unique.
  + Cependant, la variable contenance présente des valeurs qui ne suivent pas la norme attendue : la valeur **1** pour les bons cadeaux ne représente pas de problème, mais la valeur **0** indique un manque d'information et nécessite une révision.
* **Cohérence des données** :
  + Aucune valeur négative n'a été détectée.
  + Les prix d'achat sont inférieurs aux prix de vente, ce qui est conforme aux attentes commerciales.
  + Toutefois, 32 lignes montrent des prix promotionnels supérieurs aux prix de vente réguliers.
* **Erreurs supplémentaires** :
  + Les noms des clés primaires dans les tables finance, promo, et sales ne correspondent pas à ceux mentionnés dans le dictionnaire des données.
  + Les dates d'extraction vont jusqu'en octobre 2023, contrairement à la période annoncée qui s'arrête en septembre 2023.
  + Les définitions de sales\_offer et price\_offer dans la table promo sont inversées dans le dictionnaire des données
  + **2.3 Identification des principaux problèmes ou enjeux**

Le projet de création d'un tableau de bord pour la visualisation des données présente plusieurs défis potentiels à prendre en compte pour garantir que la solution choisie réponde aux besoins de l'entreprise à long terme.

* **Performance :** La performance de la base de données SQLite dépendra fortement de la méthode choisie pour la connecter à la solution de data visualisation choisi. Une connexion directe via ODBC pourrait convenir pour des volumes de données modestes, mais montrer ses limites à mesure que les données augmentent. L'extraction en CSV pourrait entraîner une charge de travail supplémentaire pour maintenir des performances acceptables, tandis que l'utilisation d'un ETL pourrait offrir une optimisation efficace, bien que potentiellement plus coûteuse et complexe.
* **Volume de Données :** L'augmentation progressive du volume de données, avec l'ajout mensuel de nouveaux produits, prix, et promotions, est un facteur déterminant dans le choix de la solution, car elle influencera directement les performances à long terme.
* **Nombre d'Utilisateurs Simultanés** : Le nombre de personnes travaillant simultanément sur la base de données et le logiciel de data visualisation est une autre problématique à prendre en compte. Une connexion directe pourrait entraîner des ralentissements, affectant à la fois les performances de ces derniers, surtout si plusieurs utilisateurs accèdent aux mêmes données en même temps.
* **Fréquence de Mise à Jour :** Une intégration en temps réel nécessiterait une infrastructure solide pour gérer les flux de données continus, tandis qu'une mise à jour quotidienne ou hebdomadaire pourrait être moins exigeante, mais nécessiterait une planification rigoureuse pour maintenir les données à jour.
* **Gestion des Conflits et Fiabilité :** Il est essentiel de garantir que la base de données reste fiable et sans conflits, avec des mécanismes robustes de sauvegarde et de restauration en place

**Besoin d'outils**

* + **3.1 Identification des outils nécessaires pour collecter les données**

Afin de pouvoir lier notre base de données à notre solution de data visualisation, nous avons besoin d'outils et de logiciels spécifiques, selon la méthode de connexion choisie :

* **ODBC Driver et ODBC Data Source Administrator (DSN) :** Permet une connexion directe et continue entre la base de données SQLite et interface de visualisation de données, facilitant l'accès en temps réel aux données.
* **Exportation en CSV :** Utilisé pour l'extraction manuelle des données depuis SQLite. Les fichiers CSV peuvent ensuite être importés dans le logiciel de datavisualisation. Pour l'automatisation, un script en Python peut être utilisé.
* **Outils ETL (Power Query, Knime) :**
* Power Query est intégré à Power BI et est idéal pour des transformations de données simples à modérées avant la visualisation.
* Knime est un outil ETL externe, plus puissant et flexible, adapté aux transformations complexes et à des workflows plus avancés, mais nécessite une licence.
* **Data Lake (ex. Azure Data Lake, Amazon S3) :** Utilisé pour centraliser et stocker les données à grande échelle avant leur intégration dans l’outil de dataviz. Cela convient aux solutions nécessitant la gestion de volumes importants de données non structurées ou semi-structurées.
* **Power BI Dataflows :** Permet de créer des flux de données réutilisables, facilitant l'intégration et la gestion de données provenant de plusieurs sources dans Power BI.
* **Scripts Python ou autres langages similaires :** Utilisés pour automatiser l'extraction en CSV et la mise à jour des données dans la solution de data visualisation choisie.
* **3.2 Identification des outils nécessaires pour traiter les données**

Afin de pouvoir utiliser correctement les données dans le logiciel de visualisation des données, il est nécessaire de disposer d'outils de transformation des données, notamment pour le nettoyage et la modification. Voici les options disponibles :

* **SGBD-R**: Déjà disponible puisque notre base de données est SQLite, le SQL permet de réaliser des transformations directement au sein de la base de données.
* **Excel** : Utilisé pour la modification des fichiers CSV, Excel convient aux modifications simples sur des volumes de données limités.
* **Outils ETL** : Power Query, intégré à Power BI, est idéal pour des transformations de données courantes. Pour des transformations plus complexes, d'autres outils comme KNIME, SSIS, ou Alteryx peuvent être utilisés.
* **Langages de programmation** : Des langages comme **Python** ou **R** permettent des transformations avancées, offrant une grande flexibilité pour le traitement des données.
  + **3.3 Identification des outils nécessaires pour analyser les données**

L'outil choisi pour analyser les données doit permettre de créer des visualisations diversifiées, agréables, et user-friendly, adaptées à un usage quotidien. Il doit également offrir la possibilité de filtrer les données par marché ou produit pour répondre aux besoins spécifiques des équipes.

Power BI est particulièrement adapté, car il combine des capacités de visualisation puissantes avec une interface intuitive et conviviale, facilitant l'utilisation par des utilisateurs non techniques. Tableau est une autre option robuste pour les visualisations.

En revanche, des outils comme Python et R permettent des analyses avancées, mais leur utilisation demande des compétences en programmation, ce qui les rend moins accessibles pour les équipes.

**Solutions d'extraction, de traitement et de visualisation**

* + **4.1 Solutions proposées pour l'extraction des données**

L'extraction des données de la base SQLite vers la solution de visualisation choisi peut se faire de différentes manières, chacune ayant ses avantages et inconvénients. Voici les principales options :

* Connexion directe à SQLite via ODBC
  + Avantages :
    - Accès en temps réel : Les données sont extraites directement de la base SQLite, garantissant que les analyses sont basées sur les informations les plus récentes.
    - Aucune duplication : Les données ne sont pas dupliquées, ce qui réduit les risques de divergence entre les versions.
  + Inconvénients :
    - Performance limitée : La performance peut être affectée par le volume de données et le nombre d'utilisateurs simultanés, en particulier si la base de données SQLite devient volumineuse.
    - Configuration : Nécessite la configuration d'un ODBC driver et d'un DSN (Data Source Name), ce qui peut être technique et nécessite un certaine compétence.
* Extraction en CSV
  + Avantages :
    - Simplicité : Exporter les données en CSV est simple et ne nécessite pas de configuration complexe.
    - Portabilité : Les fichiers CSV peuvent être facilement déplacés, partagés, et importés dans l’outil de dataviz.
  + Inconvénients :
    - Manque de mise à jour en temps réel : Une fois exportées, les données sont statiques. Pour obtenir les données les plus récentes, une nouvelle extraction est nécessaire.
    - Processus manuel : Si l'extraction n'est pas automatisée, cela peut devenir fastidieux et sujet à des erreurs humaines.
* Utilisation d'un ETL
  + Avantages :
    - Automatisation : Un ETL (Extract, Transform, Load) comme KNIME peut automatiser le processus d'extraction, réduisant ainsi le besoin d'intervention manuelle.
    - Flexibilité : Les outils ETL permettent de configurer des extractions complexes, combinant des données de plusieurs sources avant de les importer dans l’application de visualisation des données.
  + Inconvénients :
    - Coût et Complexité : Certains ETL nécessitent des licences payantes et peuvent être complexes à configurer, surtout pour des utilisateurs non techniques.
    - Temps d'exécution : L'extraction via un ETL peut prendre plus de temps que d'autres méthodes, en particulier pour des volumes de données importants.
  + **4.2 Solutions proposées pour le traitement des données**

Le traitement et la transformation des données sont essentiels pour préparer des informations précises et exploitables avant leur visualisation. Voici les options disponibles pour réaliser ces transformations, avec leurs avantages et inconvénients :

* SQL/Transformation dans la base de données (Connexion directe)
  + Avantages :
    - Traitement en temps réel : Permet de réaliser les transformations directement dans la BDD, ce qui évite de devoir exporter et réimporter les données.
  + Inconvénients :
    - Capacités limitées : SQLite est moins performant pour les transformations complexes ou volumineuses.
    - Temps de traitement : Les transformations complexes en SQL peuvent être chronophages, nécessitant un entretien régulier.
    - Maintenabilité : Les nouvelles saisies doivent respecter les mêmes règles que les transformations appliquées, sinon elles devront être réappliquées, augmentant la charge de travail. Nécessite également la mise a jour du dictionnaire de données et du schéma relationnel.
* Fichiers CSV/Modification via Excel ou Power Query
  + Avantages :
    - Flexibilité pour traitement externe : Les fichiers CSV peuvent être manipulés facilement dans Excel ou Power Query, facilitant le nettoyage et les transformations.
    - Simplicité : Idéal pour des transformations simples et la création de colonnes dérivées à partir des données existantes.
    - Intégration dans Power BI : Power Query, intégré à Power BI, simplifie le traitement des données extraites.
  + Inconvénients :
    - Processus manuel : L'extraction et la manipulation des fichiers CSV peuvent devenir fastidieuses si les données doivent être fréquemment mises à jour.
    - Performances limitées : Si la structure des données change, il faudra reconfigurer les transformations dans Power Query, ce qui peut affecter les performances, surtout avec des volumes de données importants ou plusieurs utilisateurs simultanés.
* Outils ETL (Power Query et KNIME)
  + Power Query (Utilisation avec fichiers CSV et connexion directe)
    - Avantages :
      * Intégré à Power BI : Directement accessible depuis Power BI, il est facile à utiliser pour les utilisateurs non techniques.
      * Transformation courante des données : Efficace pour nettoyer, ajouter des colonnes, filtrer, et supprimer des lignes dans des fichiers CSV ou via une connexion directe à la base de données.
    - Inconvénients :
      * Limitations pour les transformations complexes : Power Query peut montrer ses limites pour les manipulations plus sophistiquées ou volumineuses.
      * Performances dépendantes de la structure : La performance peut être affectée si la structure de la base de données change ou si elle est utilisée par plusieurs utilisateurs simultanément.
  + KNIME (connexion directe à la BDD)
    - Avantages :
      * Flexibilité et puissance : Adapté aux transformations complexes, avec la possibilité de créer des workflows avancés et des transformations en temps réel sur les données issues de la base de données.
      * Intégration avec Power BI : Les données transformées peuvent être immédiatement utilisées dans Power BI, avec une mise à jour automatique dès que les données de la base changent.
    - Inconvénients :
      * Courbe d'apprentissage : Plus difficile à maîtriser, nécessitant souvent une formation spécifique pour les utilisateurs non techniques.
      * Complexité et coût : La configuration de workflows complexes peut être longue et certaines fonctionnalités avancées nécessitent une licence payante.
* Langages de programmation (Python, R)
  + Avantages :
    - Puissance et flexibilité : Permet de réaliser des transformations complexes, nettoyer les données, et créer des colonnes dérivées à partir de calculs avancés, le tout directement relié à la base de données.
    - Large gamme de bibliothèques : Une vaste sélection de bibliothèques est disponible pour simplifier et automatiser le traitement des données.
  + Inconvénients :
    - Nécessite des compétences en programmation : Ces langages demandent une expertise technique, ce qui peut compliquer leur adoption par des membres de l'équipe non techniques.
    - Complexité : Le développement et la maintenance des scripts peuvent être complexes et nécessiter un temps de développement supplémentaire.
  + **4.3 Solutions proposées pour la visualisation des données**

La visualisation des données est une étape clé pour rendre les informations accessibles et exploitables par les utilisateurs finaux. Il est essentiel de choisir un outil qui permette de créer des visualisations claires, diversifiées et interactives, tout en étant simple d'utilisation. Voici les principales options disponibles :

* Power BI
  + Avantages :
    - Référence du marché : Power BI est l'outil de visualisation de données le plus connu et le plus utilisé, développé par Microsoft. Il offre une large gamme de fonctionnalités pour analyser les données et optimiser le processus décisionnel.
    - Flexibilité de connexion : Power BI offre la possibilité de se connecter aux données de différentes manières, que ce soit en direct sur la base de données, via l'importation de fichiers CSV, ou en utilisant un ETL externe comme KNIME. Cette flexibilité permet d'adapter l'intégration des données en fonction des besoins spécifiques de l'entreprise et des ressources disponibles.
    - Sécurité et gouvernance : Grâce à Microsoft Cloud App Security, Power BI offre une gestion fluide et sécurisée des données.
    - Intelligence Artificielle : Power BI intègre des fonctionnalités d'IA pour optimiser le traitement des données.
    - User-friendly : Son interface est intuitive, ce qui le rend accessible même pour des utilisateurs non techniques.
  + Inconvénients :
    - Coût : Bien que Power BI propose une version gratuite, les fonctionnalités avancées nécessitent une licence payante.
    - Performance : Sur des ensembles de données très volumineux, Power BI peut parfois montrer des limites en termes de performance.
* Tableau
  + Avantages :
    - Simplicité d'utilisation : Tableau est particulièrement adapté pour les débutants en data visualisation, permettant de créer des visualisations efficaces sans nécessiter de compétences en programmation.
    - Rapidité d'analyse : Tableau offre une analyse rapide des données, même pour de grandes quantités.
    - Flexibilité de connexion : Tableau permet de se connecter à une vaste gamme de sources de données, incluant des bases de données relationnelles, des fichiers CSV, des feuilles de calcul Excel, ainsi que via des ETL externes. Cette flexibilité permet aux utilisateurs de travailler avec diverses sources de données en fonction de leurs besoins et de maintenir des données à jour en temps réel
    - Partage facile : Les visualisations peuvent être partagées en quelques clics sur PC ou mobile, ce qui facilite la collaboration.
  + Inconvénients :
    - Coût élevé : Tableau est généralement plus coûteux que Power BI, surtout pour les entreprises ayant besoin de licences multiples.
    - Apprentissage nécessaire : Bien que simple à utiliser, maîtriser toutes les fonctionnalités de Tableau peut nécessiter du temps et une formation.
* Qlik
  + Avantages :
    - Analyse guidée : Qlik offre des tableaux de bord prédéfinis et des analyses associatives, ce qui aide les utilisateurs à explorer les données de manière intuitive.
    - Performances : Le moteur de recherche avancé de Qlik permet une extraction rapide des données sans besoin de mettre en place une base de données.
    - Flexibilité de connexion : Qlik offre la possibilité de se connecter à diverses sources de données, incluant des bases de données relationnelles, des fichiers CSV, des feuilles de calcul, et des flux de données en temps réel. Grâce à des outils comme Qlik Data Integration, il permet également l'intégration de données en continu, assurant que les analyses sont basées sur les informations les plus récentes, indépendamment de la source.
    - Optimisation des ressources : Les données sont stockées sur la RAM du serveur, ce qui accélère les requêtes et réduit la taille des fichiers.
  + Inconvénients :
    - Complexité pour les débutants : Qlik peut être moins intuitif que Power BI ou Tableau pour les nouveaux utilisateurs.
    - Coût : Comme les autres outils premium, Qlik nécessite une licence, ce qui peut représenter un investissement important.
* Looker Studio (ex Google Data Studio)
  + Avantages :
    - Gratuit : Looker Studio est un outil de data visualisation entièrement gratuit, ce qui le rend accessible pour les petites structures ou les budgets limités.
    - Simplicité et collaboration : Facile à utiliser et entièrement basé sur le Cloud, Looker Studio permet une collaboration en temps réel avec des fonctionnalités similaires à Google Docs.
    - Centralisation des données : Looker Studio facilite la récupération et la centralisation des données provenant de multiples sources web, y compris Google Analytics, réseaux sociaux, et autres.
  + Inconvénients :
    - Fonctionnalités limitées : Bien que performant pour des visualisations simples, Looker Studio est moins puissant que Power BI ou Tableau pour des analyses complexes.
    - Dépendance à l'écosystème Google : Looker Studio est particulièrement optimisé pour les utilisateurs déjà intégrés dans l'écosystème Google, ce qui peut limiter son utilisation pour d'autres types de sources de données.

**Cohérence des solutions avec le besoin**

* + **5.1 Explication détaillée de chaque solution proposée**

A la suite du développement présenté dans le point 4.3, nous avons choisi Power BI comme solution de data visualisation pour ce projet. Cette décision a été prise en raison de ses nombreux avantages, notamment sa flexibilité, son intégration avec d'autres outils, et sa convivialité pour les utilisateurs finaux. Dans cette section, nous détaillons les trois principales méthodes pour intégrer les données dans Power BI : la connexion directe à la base de données SQLite, l'extraction des données en CSV, et l'utilisation d'un ETL, en particulier KNIME.

1. Connexion directe à la base de données SQLite
   * Étapes de mise en œuvre :
     + Installer et configurer un ODBC driver pour SQLite.
     + Créer un Data Source Name (DSN) pour établir la connexion entre SQLite et Power BI.
     + Configurer les requêtes dans Power BI pour extraire les données directement depuis la base de données.
   * Avantages :
     + Accès en temps réel : Les données sont toujours à jour dans Power BI, sans besoin d'extraction manuelle.
     + Simplicité : La connexion directe réduit le besoin de manipulations supplémentaires des données.
   * Inconvénients :
     + Performance limitée : Peut-être moins performant avec des volumes de données élevés ou plusieurs utilisateurs simultanés.
     + Capacité limitée pour les transformations complexes : Les transformations en SQL peuvent être complexes et chronophages.
2. Extraction des données en CSV
   * Étapes de mise en œuvre :
     + Extraire les données de SQLite au format CSV.
     + Importer les fichiers CSV dans Power BI.
     + Utiliser Power Query pour transformer les données au besoin avant de les visualiser.
   * Avantages :
     + Simplicité : Facile à mettre en place et à utiliser, idéal pour des transformations de données simples.
     + Flexibilité : Les fichiers CSV peuvent être manipulés avec divers outils, comme Excel.
   * Inconvénients :
     + Pas de mise à jour en temps réel : Les données doivent être extraites manuellement ou via un script à chaque mise à jour.
     + Processus manuel : La manipulation des fichiers CSV peut devenir fastidieuse si elle n'est pas automatisée.
3. Utilisation d’un ETL (KNIME)
   * Étapes de mise en œuvre :
     + Configurer KNIME pour se connecter à la base de données SQLite.
     + Créer des workflows de transformation des données dans KNIME.
     + Exporter les données transformées directement vers Power BI via l'intégration de KNIME.
   * Avantages :
     + Puissance et flexibilité : KNIME permet des transformations complexes et l'intégration de multiples sources de données.
     + Mise à jour en temps réel : Les données peuvent être transformées en temps réel et immédiatement disponibles dans Power BI.
   * Inconvénients :
     + Complexité : KNIME peut nécessiter une courbe d'apprentissage plus importante, surtout pour les utilisateurs non techniques.
     + Coût potentiel : Certaines fonctionnalités avancées peuvent nécessiter une licence payante.
   * **5.2 Alignement des solutions avec les besoins identifiés**

Dans cette section, nous examinons comment chaque solution proposée répond aux enjeux spécifiques de l'entreprise, notamment le suivi de l'évolution des prix, l'analyse des performances promotionnelles, la mise à jour des données, la facilité d'utilisation, et la performance.

* Connexion directe à la base de données SQLite
  + Alignement avec les besoins :
    - Mise à jour des données : Cette solution permet une mise à jour en temps réel, ce qui est crucial pour suivre l'évolution des prix et analyser les performances promotionnelles sans délai.
    - Facilité d'utilisation : Pour les utilisateurs finaux, la connexion directe intégrée à Power BI offre une interface simple et directe, bien que la configuration initiale nécessite une expertise technique.
    - Performance : Bien que la mise à jour en temps réel soit un atout, cette solution peut rencontrer des limitations en termes de performance lorsque les volumes de données augmentent ou lorsque plusieurs utilisateurs accèdent simultanément à la base de données.
* Extraction des données en CSV
  + Alignement avec les besoins :
    - Mise à jour des données : L'extraction en CSV ne permet pas une mise à jour en temps réel, ce qui peut limiter la pertinence des analyses pour des suivis dynamiques comme l'évolution des prix ou les promotions.
    - Facilité d'utilisation : Les fichiers CSV peuvent être facilement manipulés et importés dans Power BI, mais l'extraction manuelle peut devenir fastidieuse et sujette à des erreurs.
    - Performance : Cette méthode est simple à mettre en place, mais elle n'est pas adaptée pour gérer des mises à jour fréquentes ou des volumes de données importants, ce qui peut nuire à la réactivité des analyses.
* Utilisation d’un ETL (KNIME)
  + Alignement avec les besoins :
    - Mise à jour des données : KNIME permet de transformer les données en temps réel et de les intégrer immédiatement dans Power BI, assurant ainsi que les analyses, telles que le suivi des prix et l'évaluation des promotions, soient toujours basées sur les données les plus récentes.
    - Facilité d'utilisation : Bien que KNIME offre une grande flexibilité et puissance, sa courbe d'apprentissage peut être un obstacle pour les utilisateurs non techniques. Cependant, une fois maîtrisé, il offre une solution robuste pour des analyses complexes.
    - Performance : KNIME est capable de gérer des transformations complexes et des volumes de données importants sans sacrifier la performance, ce qui en fait une solution idéale pour répondre aux besoins de l'entreprise en matière d'analyse détaillée et en temps réel.
  + **5.3 Avantages et limites de chaque approche**
* Connexion directe à la base de données SQLite
  + Avantages :
    - Mise à jour en temps réel : Permet un accès immédiat aux données les plus récentes, ce qui est idéal pour les pages récapitulatives dynamiques et les analyses en temps réel.
    - Simplicité d'accès : Les chefs de produits peuvent facilement accéder aux analyses à jour sans nécessiter d'extractions ou de transformations manuelles.
  + Limites :
    - Performance limitée : Peut devenir moins réactive avec de gros volumes de données ou plusieurs utilisateurs simultanés, ce qui pourrait affecter la fluidité de l'expérience utilisateur.
* Extraction des données en CSV
  + Avantages :
    - Simplicité de mise en œuvre : Facile à configurer, avec des outils familiers comme Excel pour les transformations.
    - Flexibilité pour des transformations simples : Idéal pour des analyses basiques ou des données qui ne changent pas fréquemment.
  + Limites :
    - Pas de mise à jour en temps réel : Nécessite une extraction manuelle ou scriptée des données, ce qui peut rendre les pages récapitulatives obsolètes et limiter l'accès rapide aux analyses pour les chefs de produits.
* Utilisation d’un ETL (KNIME)
  + Avantages :
    - Puissance et flexibilité : Capable de gérer des transformations complexes et de mettre à jour les données en temps réel, ce qui est crucial pour des analyses détaillées et des pages récapitulatives dynamiques.
    - Intégration directe avec Power BI : Permet une automatisation complète du flux de données, offrant ainsi un accès facile et immédiat aux analyses pour les chefs de produits.
  + Limites :
    - Complexité : Plus difficile à maîtriser pour les utilisateurs non techniques, et potentiellement coûteux en termes de licences pour des fonctionnalités avancées.

**Conclusion**

* + **Synthèse des principales conclusions de l'analyse**

L'analyse a permis d'identifier trois méthodes principales pour intégrer les données dans Power BI : la connexion directe à SQLite, l'extraction en CSV, et l'utilisation d'un ETL comme KNIME. Chacune de ces approches présentent des avantages distincts : la connexion directe offre une mise à jour en temps réel, l'extraction en CSV est simple mais moins réactive, et KNIME combine flexibilité et puissance pour des transformations complexes. Cependant, les limites en termes de performance, de complexité, et de coût varient, et doivent être considérées en fonction des besoins spécifiques de l'entreprise.

* + **Recommandations pour la mise en œuvre des solutions proposées**

Au vu des besoins spécifiques du projet et de l'analyse réalisée, l'utilisation de KNIME semble être la solution la plus adaptée. Voici pourquoi :

* Puissance et flexibilité : KNIME est capable de gérer des volumes de données importants et de réaliser des transformations complexes. Cette capacité est essentielle pour l'analyse approfondie des performances promotionnelles et le suivi des marges en temps réel.
* Intégration directe : KNIME peut se connecter directement à la base de données SQLite pour recevoir les données brutes, les transformer selon les besoins, puis les transmettre à Power BI pour la visualisation. Cette intégration permet de maintenir les données à jour en temps réel, un aspect crucial pour l'entreprise.
* Adaptabilité : KNIME offre la possibilité d'intégrer plusieurs sources de données, ce qui est un atout pour l'avenir. Par exemple, en ajoutant des données CRM, il serait possible de travailler sur la segmentation client et de développer des promotions ciblées, renforçant ainsi l'analyse stratégique de l'entreprise.

En somme, KNIME n'est pas seulement une solution qui répond aux besoins actuels, mais c'est aussi un outil flexible et évolutif, capable de s'adapter aux futurs besoins de l'entreprise. Son utilisation permettra non seulement de visualiser les données efficacement dans Power BI, mais aussi de soutenir la croissance et l'adaptation continue de l'entreprise dans un environnement en constante évolution.

Ainsi, je recommande l'adoption de KNIME pour ce projet, en raison de sa robustesse, de sa flexibilité, et de son potentiel à répondre aux exigences présentes et futures de l'entreprise.

**[Annexes]**

* **Liste des sources de données utilisées**

[**https://learn.microsoft.com/fr-fr/power-bi/connect-data/desktop-data-sources**](https://learn.microsoft.com/fr-fr/power-bi/connect-data/desktop-data-sources)

[**https://docs.knime.com/2020-07/powerbi\_integration\_user\_guide/index.html#introduction**](https://docs.knime.com/2020-07/powerbi_integration_user_guide/index.html#introduction)

[**https://learnsql.fr/blog/sql-et-power-bi-transformer-les-donnees-en-connaissances/**](https://learnsql.fr/blog/sql-et-power-bi-transformer-les-donnees-en-connaissances/)

[**https://forum.knime.com/t/knime-vs-powerquery/45311**](https://forum.knime.com/t/knime-vs-powerquery/45311)

[**https://www.lepont-learning.com/fr/top-outils-visualisation-donnees-dataviz/**](https://www.lepont-learning.com/fr/top-outils-visualisation-donnees-dataviz/)

[**https://python.doctor/page-database-data-base-donnees-query-sql-mysql-postgre-sqlite**](https://python.doctor/page-database-data-base-donnees-query-sql-mysql-postgre-sqlite)

* **Description détaillée des outils recommandés**

**SQLite :**

**https://www.sqlite.org/**

**Knime :**

**https://www.knime.com/  
PowerBI :**

[**https://www.microsoft.com/fr-fr/power-platform/products/power-bi**](https://www.microsoft.com/fr-fr/power-platform/products/power-bi)

* **Exemples de visualisations de données proposées**

Une image contenant texte, capture d’écran, Police, logiciel

Description générée automatiquement

Une image contenant texte, capture d’écran, logiciel, carte

Description générée automatiquement