



Rapport d'Analyse

Date : 26/03/2025

1. Introduction

Objectif du rapport :

L'objectif de ce rapport est d'analyser les différentes solutions d'extraction, de traitement et de visualisation des données disponibles pour Bottleneck, la boutique spécialisée dans la vente de vin en ligne. Ce rapport vise à fournir une analyse détaillée des options envisageables pour répondre aux besoins de l'entreprise tout en tenant compte de la structure actuelle des données.

2. Contexte de l'analyse

2.1 Description de la situation actuelle :

Bottleneck a récemment mis à jour ses bases de données, couvrant la période du 1er octobre 2022 au 30 septembre 2023. Ces bases contiennent des informations variées liées aux ventes, aux stocks, et aux comportements des clients. L'entreprise souhaite désormais rendre ces données accessibles via des outils de data visualisation pour une analyse plus approfondie. La gestion des stocks a déjà été optimisée, mais le besoin de mise en place d'une solution de data visualisation devient une priorité pour faciliter la prise de décision.

2.2 Évaluation de la pertinence des données existantes :

Les données actuelles semblent bien structurées avec un schéma relationnel clair. Cependant, des erreurs mineures ou des données manquantes ont été identifiées lors de l'analyse précédente, bien que cela n'entrave pas leur utilisation pour la visualisation. Il est important de noter que les données couvrent une période spécifique

et sont représentatives des transactions et des stocks pendant cette période, ce qui en fait une base solide pour les futures analyses.

2.3 Identification des principaux problèmes ou enjeux :

- **Accès aux données :** La gestion des accès aux bases de données et le choix de la méthode d'extraction sont des enjeux cruciaux.
 - **Qualité des données :** Bien que les erreurs majeures aient été corrigées, il pourrait subsister des incohérences mineures à vérifier.
 - **Besoins de visualisation :** Il est nécessaire de définir les types de visualisations les plus utiles pour la prise de décision, en fonction des indicateurs clés de performance (KPI).
-

3. Besoin d'outils

3.1 Identification des outils nécessaires pour collecter les données :

Pour collecter les données depuis les bases de données existantes, plusieurs outils peuvent être utilisés :

- **Connexion directe à la base de données :** Une option idéale si les données doivent être récupérées en temps réel et si la base de données est stable et bien structurée.
- **Extraction en CSV :** Simple à mettre en œuvre, mais plus lente et peut devenir un goulot d'étranglement si les volumes de données augmentent.
- **Utilisation d'un ETL (Extract, Transform, Load) :** Outils comme PowerQuery ou Knime peuvent être utilisés pour extraire, transformer et charger les données dans un format adapté pour l'analyse.

3.2 Identification des outils nécessaires pour traiter les données :

Une fois les données collectées, plusieurs outils peuvent être utilisés pour les traiter :

- **ETL(Extract,Transform,Load) :** Permet de transformer les données en un format prêt pour l'analyse et de gérer les nettoyages ou agrégations.
- **Langages de programmation (Python/R) :** Si des traitements plus complexes sont nécessaires, Python ou R peuvent être utilisés pour effectuer des analyses statistiques ou des nettoyages avancés.

3.3 Identification des outils nécessaires pour analyser les données :

Les outils d'analyse des données doivent permettre de créer des visualisations et des rapports interactifs :

- **Power BI / Tableau** : Ces outils de BI (Business Intelligence) permettent de créer des tableaux de bord dynamiques et des visualisations interactives.
- **Google Data Studio(Looker Studio)** : Solution gratuite et facile à prendre en main, idéale pour les petites et moyennes entreprises.
- **Tableau Software** : Plus avancé, offrant des capacités de visualisation de données complexes et des intégrations étendues.

4. Solutions d'extraction, de traitement et de visualisation

4.1 Solutions proposées pour l'extraction des données :

- **Connexion directe à la base de données** : Cette méthode permet d'obtenir des données en temps réel, avec un minimum de latence. Elle est idéale pour des rapports en continu ou des mises à jour fréquentes des données.
- **Extraction en CSV** : Utile pour des exports ponctuels ou pour des analyses non critiques. Cependant, cette méthode peut devenir moins efficace si les volumes de données augmentent.
- **ETL avec PowerQuery ou Knime** : Cette approche est idéale si l'on souhaite nettoyer et transformer les données avant leur analyse. Elle permet également de programmer des extraits réguliers.

Notre choix :

Pour collecter les données depuis la base SQLite et les transférer vers l'outil de data visualisation, nous avons opté pour ODBC (Open Database Connectivity).

Cette solution permet :

- ✓ Une connexion fluide et sécurisée entre la base SQLite et l'outil BI.
- ✓ Une extraction dynamique des données sans nécessiter d'export manuel.
- ✓ Une meilleure intégration que les fichiers CSV pour des mises à jour fréquentes.

Mise en œuvre :

Nous avons téléchargé et configuré le connecteur ODBC spécifique à SQLite, qui sert d'interface entre la base de données et l'outil de Business Intelligence.

4.2 Solutions proposées pour le traitement des données :

- **Utilisation d'un ETL pour le traitement des données** : PowerQuery ou Knime peuvent être utilisés pour automatiser les tâches de transformation, de nettoyage, et de préparation des données.
- **Python/R pour des traitements avancés** : Ces langages offrent plus de flexibilité pour effectuer des analyses avancées, mais nécessitent des compétences techniques plus poussées.

Notre choix :

Pour le traitement des données, nous avons retenu Power Query, l'ETL intégré à Power BI.

Avantages de ce choix :

✓ Nettoyage et transformation efficaces :

- Prépare les données pour la visualisation en éliminant les erreurs et en standardisant les formats.
- Permet des opérations telles que le filtrage, l'agrégation et la fusion de sources multiples.
-

✓ Automatisation des traitements :

- Les étapes de transformation sont enregistrées et rejouées automatiquement lors des mises à jour.
- Réduit les interventions manuelles, garantissant une plus grande fiabilité des données.

✓ Flexibilité pour les futurs besoins :

- Facilement adaptable à de nouveaux jeux de données ou règles métier.
- Permet d'enrichir les processus existants sans reprogrammation.

4.3 Solutions proposées pour la visualisation des données :

- **Power BI / Tableau** : Ces outils sont adaptés pour une visualisation en temps réel et offrent une grande variété de types de graphiques et de tableaux de bord.
- **Google Data Studio(Looker Studio)** : Recommandé pour des visualisations moins complexes ou pour des utilisateurs n'ayant pas de compétences en BI.

- **Tableau Software** : Une solution plus avancée pour des besoins de visualisation complexes, permettant des analyses plus détaillées.
- **Plotly (l'outil de dataviz des développeurs Python):**

Notre choix :

Nous avons retenu Power BI pour les raisons suivantes :

- ✓ Intégration native avec notre outil ETL (Power Query)
- ✓ Partie Desktop gratuite (open source), permettant de créer des dashboards professionnels sans coût supplémentaire
- ✓ Grande variété de visualisations et une interactivité poussée
- ✓ Facilité de partage et de collaboration via Power BI Service
- ✓ Adapté aussi bien aux débutants qu'aux experts en BI

Cette solution offre le meilleur compromis entre puissance, simplicité d'utilisation et coût, répondant parfaitement à nos besoins en matière de reporting et d'analyse de données.

5. Cohérence des solutions avec le besoin

5.1 Explication détaillée de chaque solution proposée :

Chaque solution a été choisie en fonction des besoins spécifiques de Bottleneck :

- **Connexion directe à la base de données** : Solution idéale pour les entreprises ayant besoin de mises à jour en temps réel et de rapports dynamiques.
- **ETL (PowerQuery)**: Approprié si des transformations ou nettoyages de données sont nécessaires avant l'analyse.
- **Power BI** : Outils puissants pour la création de rapports dynamiques et d'analyses visuelles interactives.

5.2 Alignement des solutions avec les besoins identifiés :

Les solutions proposées sont parfaitement alignées avec les besoins de Bottleneck, notamment la nécessité d'obtenir des rapports et visualisations interactifs basés sur des données actualisées.

5.3 Avantages et limites de chaque approche :

- **Connexion directe** : Avantage : données en temps réel, pas de duplication.
Limite : nécessite une gestion stricte des accès et peut-être lent si la base est volumineuse, risque de surcharge.
- **ETL** : Avantage : nettoyage et transformation automatisés. Limite : Nécessite une configuration ETL (plus complexe à mettre en place).
- **Power BI** : Avantage : visualisations riches et interactives. Limite : coût élevé (La version gratuite (Power BI Desktop) a des limites en termes de taille des jeux de données (1 Go en importation), nécessite une certaine expertise (peut nécessiter des licences payantes).

6. Conclusion

Synthèse des principales conclusions de l'analyse :

En fonction des besoins de Bottleneck, la meilleure solution semble être l'utilisation d'une combinaison de la connexion directe à la base de données pour l'extraction, accompagnée d'un ETL pour le traitement des données, et Power BI pour la visualisation. Cela permettra à l'entreprise de disposer de données actualisées en temps réel et de visualisations interactives pour la prise de décision.

Recommandations pour la mise en œuvre des solutions proposées :

- Mise en place d'une connexion sécurisée à la base de données (ODBC) :

Utilisation d'ODBC : Configurer une connexion fiable et sécurisée entre la base de données (SQLite) et Power BI.

- Choix de l'outil ETL adapté aux besoins

PowerQuery (intégré à Power BI) : Idéal pour nettoyer, transformer et préparer les données avant analyse. Et aussi pour Automatisation des flux de données pour réduire les interventions manuelles.

- Formation des utilisateurs à

Pour permettre aux équipes de créer et exploiter des tableaux de bord de manière autonome.

Options de formation :

- ✓ Formations certifiantes (payantes) proposées par des centres agréés (Microsoft Learning Partners).
- ✓ Formations en ligne (MOOC, YouTube, Microsoft Learn) pour une approche plus flexible.
- ✓ Ateliers internes (si compétences disponibles en interne).

[Annexes]

- **Liste des sources de données utilisées** : Base de données SQLite, schéma relationnel, dictionnaire de données.
- **Description détaillée des outils recommandés** : Power BI, Tableau, PowerQuery, Knime.
- **Exemples de visualisations de données proposées** : Graphiques des ventes, cartes de chaleur des stocks, diagrammes des tendances.