

Rapport d'Analyse solutions de data visualisation

Date: 25/10/2024

## Introduction

* + Objectif du rapport
* Ce rapport a pour objectif d’analyser les différentes solutions de data visualisation et identifier la plus pertinente pour traiter les données de la base de données nouvellement créée en interne. J’expliquerai donc les avantages et les inconvénients de chaque solution avant de vous donner mes recommandations.
  + Contexte de l'analyse
* Après une refonte de leur base de données pour en simplifier l’accès et le traitement nous souhaitons maintenant user de solutions de data visualisation pour traiter efficacement les données et prendre des décisions éclairées plus rapidement. Plusieurs choix s’offrent à nous : Une connexion directe à la base de données ; l’extraction des fichiers au format CSV ; l’utilisation d’un logiciel ETL (Extract, Transform, Load). Trois solutions qui demandent une analyse approfondie car elle impactera l’entreprise à court, moyen et long terme.

## État des lieux et évaluation de la pertinence

* + 2.1 Description de la situation actuelle
* Maintenant que le chantier de refonte de la base de données est terminé, il faut accéder à ces données pour faire des analyses pertinentes et en temps réel afin de permettre une prise de décision efficace et rapide de la part des chefs de produits et du Directeur Général.
  + 2.2 Évaluation de la pertinence des données existantes

Les données sont maintenant structurées dans une base de données avec une structure en étoile qui facilite la lecture et l’analyse de données. Certaines de ces données sont inutiles pour nos analyses, mais nous les supprimerons pendant l’étape de transformation.

* + 2.3 Identification des principaux problèmes ou enjeux
* Le principal défi pour la data visualisation est de maintenir une connexion sécurisée, fiable et pérenne entre la base de données et les outils de data visualisation notamment avec la gestion de volumes importants de données.

## Besoin d'outils

* + 3.1 Identification des outils nécessaires pour collecter les données
* La collecte des données peut se faire de 2 façons :
* - Une connexion directe à la base de données
* - L’exportation des fichiers au format CSV.

Pour ce projet, j’ai utilisé l’extraction CSV, mais je recommande la connexion directe aux bases de données qui permettre d’avoir les données à jour sans modifier le rapport

* + 3.2 Identification des outils nécessaires pour traiter les données
* Les données doivent être nettoyées, agrégées et transformées avant visualisation. Un ETL comme Power Query est nécessaire pour faciliter cette phase (bien que faisable directement en SQL ou Python)
* J’ai utilisé Power Query pour faire les transformations sur les données pour ce projet.
  + 3.3 Identification des outils nécessaires pour analyser les données
* Pour l’analyse, et surtout les résultats d’analyse, il nous faut utiliser des outils de visualisation comme Power BI, Tableau ou QlikView. Ces outils permettent la création de tableaux de bord interactifs composés de visuels intégrés et de rapports permettant une visualisation claire des analyses.

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

* + 4.1 Solutions proposées pour l'extraction des données
* Pour l’extraction des données, j’ai opté, de mon côté pour une extraction CSV car je n’ai pas accès à la base de données pour les étapes suivantes, néanmoins, les noms de tables et de colonnes étant identiques, vous pourrez très bien connecter l’ETL à votre base de données pour automatiser l’extraction et la transformation des données
  + 4.2 Solutions proposées pour le traitement des données
* Pour le traitement des données, je l’ai effectué avec l’ETL Power Query. Outils que j’affectionne particulièrement pour son côté user friendly. Avec, nous pouvons, comme dit précédemment, supprimer les données inutiles, créer de nouvelles données à partir des données existantes afin d’approfondir les analyses et détecter les éventuelles erreurs qui se trouvent dans les données.
  + 4.3 Solutions proposées pour la visualisation des données
* Pour la visualisation, en lien avec Power Query, je vous propose donc un tableau de bord Power BI, pour les mêmes raisons (user friendly). Cet outil permet une visualisation claire des données avec beaucoup de graphiques pré-enregistrés (et la possibilité d’en récupérer d’autres avec Power BI service) et d’options de personnalisation pour le rendre d’autant plus ergonomique.

## Cohérence des solutions avec le besoin

* + 5.1 Explication détaillée de chaque solution proposée
* **Extraction des données en CSV:** Permet une portabilité plus simple des données, mais nécessite des étapes manuelles ou automatisées d'exportation régulière.
* **Connexion directe à la base de données:** Solution simple et efficace, mais pouvant entraîner des problèmes de performance en fonction de la charge sur la base et du volume des données. Ainsi qu’un risque plus grand au niveau de la sécurité.
* Les deux pouvant être utiliser avec Power Query pour le traitement et la transformation des données
  + 5.2 Alignement des solutions avec les besoins identifiés
* La connexion directe vous permettra d’avoir un visuel en temps réel sur vos données. Ce qui vous permettra de prendre les mesures nécessaires sans délai.  
    
  L’extraction CSV est quant à elle plus simple et plus efficace pour ajouter des analyses ponctuelles qui diffèrent de celles que vous aurez déjà.
  + 5.3 Avantages et limites de chaque approche
* **Connexion directe :**
* + Réel gain de temps ;
* - Impact sur la performance et la sécurité.
* **Extraction en CSV :**
* + Simple et efficace pour de petites données ;
* - Peu adapté pour des traitements réguliers.
* **ETL :**
* + Flexibilité, automatisation des processus ;
* - Nécessite des ressources techniques et une configuration initiale. Nécessite au préalable soit une connexion directe, soit une extraction CSV

## Conclusion

* + Synthèse des principales conclusions de l'analyse
* La connexion directe à la base de données est assez simple à mettre en œuvre, mais elle présente des risques en termes de performance et de sécurité. L’extraction en CSV est trop limitée pour des volumes importants. L’utilisation d’un ETL, se distingue par sa flexibilité et sa capacité à traiter efficacement des volumes importants de données tout en automatisant les processus complexes, mais nécessite une connexion à la base de données pour être réellement efficace, cette solution pallie aux problèmes de performances, mais pas aux problèmes de sécurité de la connexion directe.
  + Recommandations pour la mise en œuvre des solutions proposées
* Je vous recommande d’utiliser un ETL comme **PowerQuery** pour extraire et transformer les données avant de les visualiser dans **Power BI**. Cette approche garantit des performances optimales, tout en minimisant l'intervention humaine et en maximisant la fiabilité des processus.

[Annexes]

1. Liste des sources de données utilisées

* Microsoft.com
* Powerbi.com
* Tableau.com

1. Description détaillée des outils recommandés

* **PowerQuery :** Cet outil intégré à Power BI (et Excel) sert à extraire et transformer les données provenant de diverses sources. Il offre des fonctionnalités pour nettoyer, normaliser, fusionner et enrichir les données, permettant de :
  + - Se connecter directement à la base de données interne.
    - Créer des requêtes pour filtrer et transformer les données, en gérant des opérations comme le pivot, l’agrégation, la jointure de tables, etc.
    - Automatiser les transformations pour que chaque mise à jour des données dans Power BI intègre les derniers changements.
* **Power BI :** Cet outil de data visualisation est utilisé pour créer des rapports dynamiques et interactifs. Ses fonctionnalités incluent :
  + - L’importation directe des données transformées depuis PowerQuery.
    - La création de tableaux de bord et graphiques avancés (barres, secteurs, lignes, cartes, matrices).
    - La possibilité de définir des **KPI** et des mesures personnalisées (DAX) pour des visualisations précises et adaptées aux besoins métiers.
    - Le partage et la publication de rapports via le service Power BI pour une consultation en ligne.

1. Exemples de visualisations de données proposées

**KPI dynamiques :** Mesures de performance spécifiques aux besoins métier, comme le suivi des ventes par segment, les taux de conversion, ou l’analyse des coûts opérationnels.

**Graphiques en cascade :** Pour visualiser les contributions successives aux totaux (utile pour l'analyse de rentabilité).

**Nuage de points :** Visualise la relation entre deux variables quantitatives en plaçant chaque observation sous forme de point sur des axes X et Y.

**Histogramme :** Montre la distribution d'une variable en comptant les observations dans des intervalles de valeurs.

**Graphique en secteur :** Représente les proportions de catégories par des segments dans un cercle, chaque segment illustrant la part relative d'une catégorie.

**Cartes géographiques interactives :** Pour les données localisées géographiquement.

**Graphiques de performance :** Analyse des performances par période ou produit, avec des comparaisons entre périodes.