



04/07/2025

Workshop client 2

Styleesh



FIORONI Tristan
DATAMANIACS

Table des matières

1. Introduction	2
Contexte	2
Objectif	2
2. Méthodologie	3
2.1. Audit des données.....	3
2.2. Nettoyage & Corrections.....	3
3. Analyse des données et modélisation.....	4
3.1 Données retenues pour l'analyse	4
3.2 Données complémentaires à collecter.....	4
3.3 Schéma analytique - Modèle en étoile.....	4
4. KPIs recommandés.....	5
5. Tableau de bord - Power BI (maquette).....	5
6. Justification des graphiques	6
7. Enrichissement avec données externes	8
8. Recommandations	8
9. Coûts pour la mise en place des recommandations	9
10. Annexes	12

1. Introduction

Contexte

Styleesh est une entreprise française de vente de vêtements en ligne, présente dans 5 pays avec un catalogue de plus de 3000 produits. Bien que de nombreuses données soient collectées (ventes, trafic web, clients, etc...), elles ne sont aujourd'hui pas exploitées de manière structurée. Aucun outil de pilotage, KPI ou démarche data-driven n'est actuellement mis en place.

Objectif

Ce workshop vise à poser les fondations d'une stratégie de pilotage par la donnée via la réalisation d'un premier tableau de bord décisionnel permettant de :

- Suivre les ventes par catégories, pays et canal d'acquisition
- Comprendre les comportements clients
- Identifier les produits à fort potentiel ou en perte de vitesse

2. Méthodologie

2.1. Audit des données

products.csv

- Données complètes, sans doublons ni incohérences
- Aucune valeur manquante
- Modification typage de colonne (prix, stock) – Power BI

sales.csv

- Données complètes, mais 96% des remises sont supérieurs à 1 (valeurs qui est probablement exprimées en pourcentage). Correction recommandée : division par 100.
- Modification typage de colonne (remise) – Power BI
- Difficile de connaître qui a acheté dans les ventes

customers.csv

- Données valides et complètes
- Valeurs cohérentes pour l'âge et le sexe

website_traffic.csv

- Données valides mais uniquement sur une période de 27 Jours
- 4 sources d'acquisition: Google Ads, Facebook, Direct, Emailing
- Modification typage de colonne (taux_rebond) – Power BI

2.2. Nettoyage & Corrections

- Conversion des remises > 1 en proportion (ex: 20 devient 0.20)
- Uniformisation des formats de date
- Ajout de nouvelles sources dans des données de trafic si possible

3. Analyse des données et modélisation

3.1 Données retenues pour l'analyse

- **Produits** : id, catégorie, prix, stock
- **Ventes** : quantité, remise, date, pays, canal, produit_id
- **Clients** : id, sexe, âge, pays, date d'inscription
- **Trafic web** : date, source, visites, pages vues, taux de rebond*

3.2 Données complémentaires à collecter

- Taux de retour produit
- Marge par produit
- Taux de conversion (ajout panier → achat)
- Données qualitatives sur les produits (matière, saison, etc.)
- Localisation précise des clients (code postal)

3.3 Schéma analytique - Modèle en étoile

Fait central : Sales

- Faits : id_sale, date, quantity, discount, sale_figure, fk_id_country, fk_id_product, fk_id_customer, fk_id_channel

Dimensions :

- Products : id_product, name, category, price, stock
- Customers : id_customer, age, gender, register_date, fk_id_country
- Channels : id_channel, channel_name
- Countries : id_country, country_name
- Categories : id_category, category_name

Table à part :

- Website_traffic : date, source, visites, pages_vues, taux_rebond

PS : Je n'ai pas utilisé ce schéma pour cet exercice. Ce schéma, je le mettrai en place s'il y a énormément de données pour éviter les redondances et gagner en fluidité lors des requêtes SQL.

4. KPIs recommandés

Domaine	KPI	Calcul	Objectif
Ventes	Chiffre d'affaires	$\sum (\text{quantité} \times \text{prix} \times (1 - \text{remise}))$	Suivi global des ventes
Produits	Top produits	Classement par CA ou unités vendus	Identifier les best-sellers
Produits	Stocks dormants	Stock > x jours ou ventes nulles	Optimiser le catalogue
Clients	Panier moyen	$\sum \text{CA} / \text{nb de clients}$	Suivre le pouvoir d'achat
Marketing	Taux de conversion trafic → achat	Nb ventes / nb visites	Évaluer les campagnes
UX	Taux de rebond	Donnée fourni	Améliorer l'expérience utilisateur

5. Tableau de bord - Power BI (maquette)

Le tableau de bord comprend 5 pages principales :

1. Vue globale

- CA global, évolution mensuelle
- Top 5 produits / catégories / pays

2. Vue client

- Répartition par sexe, âge, pays
- Nouveaux vs récurrents

3. Vue marketing

- Trafic par canal
- Taux de rebond / conversion

4. Vue alertes

- Produits en rupture / trop de stock
- Baisse inhabituelle de trafic ou de ventes

5. Conjoncture (Open data)

- Nombre de pouvoir d'achat arbitrage
- Somme du pouvoir d'achat par année

6. Justification des graphiques

1. Carte de score / KPI (cartes avec grands chiffres)

Pourquoi ce choix ?

→ Permet de mettre en avant les indicateurs clés comme le chiffre d'affaires total, le nombre de produits en rupture, ou encore la moyenne des visites.

Avantages :

- Impact visuel immédiat
- Facile à lire pour les décideurs
- Se concentre sur une seule valeur, donc pas de surcharge

2. Barres horizontales (chiffre d'affaires par produit, pays, catégorie...)

Pourquoi ce choix ?

→ Adapté aux comparaisons catégorielles (produits, pays, types de clients...).

Avantages :

- Lecture naturelle (du plus grand au plus petit)
- Affiche clairement les écarts
- Peut gérer un grand nombre de catégories sans confusion

3. Courbes temporelles (chiffre d'affaires ou visites par mois/année)

Pourquoi ce choix ?

→ Pour **suivre une évolution dans le temps** (saisonnalité, tendance, anomalies).

Avantages :

- Permet de détecter des pics ou des creux
- Très bon pour repérer des tendances saisonnières ou des baisses de performance

4. Diagrammes circulaires ou donuts (par sexe, pays, tranche d'âge)

Pourquoi ce choix ?

→ Pour illustrer la répartition d'une population (clients par sexe, pays...).

Avantages :

- Donne une idée rapide des parts
- Visuellement engageant
- Adapté pour des variables qualitatives simples

5. Graphiques combinés ou gauge (trafic en baisse, VentesEnBaisse)

Pourquoi ce choix ?

→ Pour **visualiser des alertes ou seuils critiques**, comme les baisses de trafic ou de ventes.

Avantages :

- Utile pour des alertes visuelles (KPI à surveiller)
 - Met en avant des dégradations de performance
6. Immédiatement compréhensible (ex : jauge semi-circulaire)

7. Enrichissement avec données externes

Source retenue : INSEE / Eurostat

- Pouvoir d'achat par tranche d'âge ou pays
- Croisement possible avec les données client pour affiner la stratégie pricing ou ciblage

Lien contenant les données utilisées pour l'open data (.xlsx) :

[Évolution de la dépense et du pouvoir d'achat des ménages | Insee](#)

Pourquoi ?

- Cela fournit un contexte économique global
- Si le pouvoir d'achat est faible ou diminue, cela peut freiner les dépenses vestimentaires
- Visualisation des années où les dépenses vestimentaires ont dû être freiné

Autres pistes envisageables

- Google Trends : tendances de recherche par produit
- Données météo : effet saisonnier sur les ventes
- Calendrier public : influence des jours fériés, soldes, vacances

8. Recommandations

- **Structuration des données** : adopter un modèle en étoile dans un datawarehouse simplifié (BigQuery, Snowflake, etc.)
- **Culture data** : sensibiliser les équipes produit, marketing et logistique à l'usage des KPIs
- **Outils** : déployer Power BI ou Looker Studio pour des tableaux de bord interactifs
- **Collecte supplémentaire** : mettre en place des traceurs (ex : Google Tag Manager) pour suivre les paniers abandonnés ou taux de conversion
- **Analyse régulière** : créer un reporting mensuel automatisé pour le pilotage opérationnel

9. Coûts pour la mise en place des recommandations

1. Structuration des données – Mise en place d’un entrepôt de données

Pourquoi ?

Actuellement, les données sont réparties entre plusieurs fichiers (ventes, clients, trafic web...). Un entrepôt de données, structuré autour d’un modèle en étoile, permet d’unifier et de centraliser toutes les informations. Cela rend les analyses plus robustes, plus rapides et évolutives.

Pourquoi BigQuery ou Snowflake ?

- **Avantages** : solutions cloud (pas d’infrastructure à gérer), scalables, rapides, intégrables à Power BI ou Looker.
- Tarification flexible selon la volumétrie réelle, 3000 à 15000€ par an
- **Choix final** : BigQuery, car très bien intégré à l’écosystème Google (Looker Studio, Google Ads, etc.) et accessible sans trop de compétences techniques.

2. Culture data – Sensibilisation des équipes

Pourquoi ?

Un bon outil ne suffit pas : il est essentiel que les équipes produit, marketing ou logistique comprennent et utilisent les KPIs dans leurs prises de décisions. Des formations courtes mais ciblées permettent de diffuser la culture data au sein de l’entreprise.

Coût estimé : 1 000 à 5 000 € (sessions de formation ponctuelles)

Pourquoi ne pas externaliser à un cabinet ?

Cela coûte cher, et déconnecte les équipes de leur propre stratégie.

Pourquoi de la formation interne ?

- Moins coûteuse, plus concrète.
- Elle peut être ciblée sur les besoins réels : comprendre un tableau de bord, utiliser un KPI, etc.
- Elle crée une culture durable.

Choix final : formation interne rapide, animée par un data analyst ou responsable marketing.

3. Déploiement d'un outil de visualisation – Power BI ou Looker Studio

Pourquoi ?

Power BI est un outil intuitif et puissant, idéal pour créer des tableaux de bord interactifs accessibles à toutes les équipes. Looker Studio (gratuit) peut être une alternative plus légère.

Choix recommandé : Power BI (Pro) pour ses fonctionnalités avancées, avec une licence à 9,40 €/mois par utilisateur.

Coût estimé : 500 à 6 000 € par an (en fonction du nombre d'utilisateurs et si le développement est externalisé)

Outil	Avantages	Inconvénients
Power BI	Puissant, automatisé, connecteurs variés	Licence à 9,40€/mois/utilisateur
Looker Studio	Gratuit, simple	Moins puissant, plus lent sur gros volumes
Tableau	Très visuel, très puissant	Coût élevé (> 70€/mois/utilisateur)

Pourquoi Power BI ?

- Meilleur rapport qualité/prix.
- Bonne documentation, communauté active.
- Permet le rafraîchissement automatique via passerelle.
- Déjà compatible avec Excel, CSV, BigQuery, etc.

4. Collecte supplémentaire de données – Tracking comportemental

Pourquoi ?

Actuellement, Styleesh ne mesure pas les paniers abandonnés, les clics ou les taux de conversion intermédiaires. En utilisant des outils comme **Google Tag Manager**, il devient possible de suivre des comportements précis sur le site (ajouts au panier, rebond, conversions...), essentiels pour l'optimisation marketing.

Coût estimé : 1 000 à 5 000 € (implémentation et maintenance)

Pourquoi ne pas utiliser uniquement Google Analytics ?

Google Analytics ne permet pas de suivre des actions très précises comme “clic sur un bouton X” ou “ajout au panier sans achat”. Il reste limité sans configuration avancée.

Pourquoi Google Tag Manager ?

- Permet de tracer **tout comportement spécifique** sur le site sans modifier le code source.
- Gratuit.
- Compatible avec tous les outils Google (Analytics, Ads...).

Choix final : Google Tag Manager, avec des déclencheurs personnalisés (panier abandonné, clic promo, etc.)

5. Reporting automatisé – Suivi continu

Pourquoi ?

Un reporting automatisé permet aux décideurs d'avoir une vision à jour sans effort. Cela renforce la réactivité face à des anomalies (ruptures de stock, baisse de trafic, etc.) et permet un pilotage efficace.

Outils : Rafraîchissement automatique Power BI, ou intégration API depuis l'entrepôt

Coût estimé : 1 500 à 5 000 € par an

Pourquoi ne pas rester avec un export manuel dans Excel ?

- Risque d'erreur humaine.
- Pas de mise à jour en temps réel.
- Peu adapté au partage ou au monitoring quotidien.

Pourquoi Power BI automatisé ?

- Rafraîchissement quotidien ou horaire possible.
- Notifications ou alertes paramétrables.
- Centralisation des données sur un seul support visuel partagé.

10. Annexes

Modèle en étoile :

