



ÉCOLE NATIONALE DES  
SCIENCES APPLIQUÉES  
Université Abdelmalek Essaâdi



## Ecommerce Insight

Architecture BI Multi-Sources pour  
E-Commerce

Power BI • ETL • Analytics • Data Visualization

### Projet Business Intelligence

Année Universitaire 2025–2026

Dépôt GitHub

[github.com/KenzaAEK/E-Commerce-Insight](https://github.com/KenzaAEK/E-Commerce-Insight)

Réalisé par :  
ABOU-EL KASEM Kenza  
EL BAKALI Malak

Encadré par :  
Pr. BADIR Hassan  
Janvier 2026

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Contexte du Projet . . . . .	7
1.2	Objectifs du Projet . . . . .	7
1.3	Périmètre Fonctionnel . . . . .	7
<b>2</b>	<b>Architecture Technique du Système</b>	<b>9</b>
2.1	Architecture Globale . . . . .	9
2.2	Modèle Dimensionnel . . . . .	9
2.2.1	Conception du Schéma en Étoile . . . . .	9
2.2.2	Tables de Faits . . . . .	10
2.2.3	Tables de Dimensions . . . . .	10
2.2.4	Dimension Temporelle Enrichie . . . . .	10
<b>3</b>	<b>Génération et Intégration des Données</b>	<b>11</b>
3.1	Introduction et Objectifs . . . . .	11
3.2	Stratégie de Génération . . . . .	11
3.2.1	Volumes Générés . . . . .	11
3.2.2	Patterns Métier Implémentés . . . . .	12
3.2.3	Segmentation RFM . . . . .	13
3.2.4	Cohérence Statistique . . . . .	13
3.3	Gestion des Anomalies et Cas Réels . . . . .	14
3.4	Architecture Multi-Sources . . . . .	14
3.4.1	Pourquoi Plusieurs Formats ? . . . . .	14
3.4.2	Justification des Choix de Formats . . . . .	14
3.5	Exécution du Script de Génération . . . . .	15
3.5.1	Lancement et Logs d'Exécution . . . . .	15
3.5.2	Vérification de la Structure des Fichiers Générés . . . . .	15
3.6	Chargement des Données dans MySQL . . . . .	16
3.6.1	Exécution du Script d'Upload SQL . . . . .	16
3.6.2	Validation dans phpMyAdmin . . . . .	17
<b>4</b>	<b>Processus ETL et Intégration</b>	<b>18</b>
4.1	Introduction . . . . .	18
4.2	Architecture Technique et Connectivité . . . . .	18
4.2.1	Environnement technique . . . . .	18
4.2.2	Choix et configuration de la connectivité ODBC . . . . .	18
4.2.3	Création du DSN et connexion à Power BI . . . . .	19
4.3	Extraction et Acquisition Multi-sources . . . . .	21
4.3.1	Extraction relationnelle (SQL / MySQL) . . . . .	21

4.3.2	Intégration des données JSON (Trafic Web) . . . . .	21
4.3.3	Intégration des données XML . . . . .	23
4.4	Transformation et Qualité des Données . . . . .	23
4.4.1	Nettoyage et standardisation . . . . .	23
4.4.2	Gestion des valeurs manquantes . . . . .	24
4.4.3	Contrôle de l'intégrité référentielle . . . . .	25
4.5	Enrichissement et Calculs ETL . . . . .	26
4.5.1	Colonnes dérivées . . . . .	26
4.6	Modélisation et Schéma Relationnel . . . . .	27
4.6.1	Schéma en étoile . . . . .	27
4.6.2	Gestion des relations . . . . .	27
<b>5</b>	<b>Mesures DAX et Intelligence Analytique</b>	<b>29</b>
5.1	Architecture des Mesures . . . . .	29
5.2	Mesures de Base . . . . .	29
5.2.1	Indicateurs de Ventes . . . . .	29
5.2.2	Indicateurs de Rentabilité . . . . .	30
5.3	Time Intelligence : Analyses Temporelles . . . . .	30
5.3.1	Comparaisons Année sur Année (YoY) . . . . .	30
5.3.2	Comparaisons Mois sur Mois . . . . .	30
5.3.3	Cumuls Année en Cours (YTD) . . . . .	31
5.3.4	Moyennes Mobiles . . . . .	31
5.4	Analyses Clients . . . . .	31
5.4.1	Indicateurs de Base . . . . .	31
5.4.2	Taux de Rétention . . . . .	32
5.5	Analyse Produits : Ranking et ABC . . . . .	32
5.5.1	Classement des Produits . . . . .	32
5.5.2	Identification du Produit Leader . . . . .	32
5.5.3	Contribution au Chiffre d'Affaires . . . . .	33
5.5.4	Nombre de Produits Vendus . . . . .	33
5.5.5	Chiffre d'Affaires Moyen par Produit . . . . .	33
5.5.6	Classification ABC (Loi de Pareto) . . . . .	33
5.6	Gestion des Retours . . . . .	34
5.7	Trafic Web et Conversion . . . . .	35
5.7.1	Indicateurs de Performance Web . . . . .	35
5.7.2	Abandon de Panier . . . . .	35
5.8	Limitations et Bonnes Pratiques . . . . .	35
5.8.1	Mesures Incompatibles avec les Filtres . . . . .	35
5.8.2	Optimisation des Performances . . . . .	36
5.9	Synthèse des Mesures Clés . . . . .	36
<b>6</b>	<b>Dashboards et Visualisations</b>	<b>38</b>
6.1	Architecture des Tableaux de Bord . . . . .	38
6.2	Dashboard 1 : Vue Exécutive . . . . .	38
6.2.1	Objectif . . . . .	38
6.2.2	Composantes Principales . . . . .	39
6.2.3	Cas d'Usage . . . . .	39
6.2.4	Valeur Ajoutée . . . . .	39
6.3	Dashboard 2 : Analyse Clients & Fidélité . . . . .	39

6.3.1	Objectif . . . . .	39
6.3.2	Composantes Principales . . . . .	40
6.3.3	Cas d'Usage . . . . .	40
6.3.4	Valeur Ajoutée . . . . .	40
6.4	Dashboard 3 : Performance Produits . . . . .	41
6.4.1	Objectif . . . . .	41
6.4.2	Composantes Principales . . . . .	41
6.4.3	Cas d'Usage . . . . .	41
6.4.4	Valeur Ajoutée . . . . .	42
6.5	Dashboard 4 : Retours & Logistique . . . . .	42
6.5.1	Objectif . . . . .	42
6.5.2	Composantes Principales . . . . .	42
6.5.3	Cas d'Usage . . . . .	43
6.5.4	Valeur Ajoutée . . . . .	43
6.6	Dashboard 5 : Trafic Web & Conversion . . . . .	43
6.6.1	Objectif . . . . .	43
6.6.2	Composantes Principales . . . . .	43
6.6.3	Cas d'Usage . . . . .	44
6.6.4	Valeur Ajoutée . . . . .	44
6.7	Interactivité Globale . . . . .	44
6.7.1	Fonctionnalités Communes . . . . .	44
	<b>Conclusion Générale</b>	<b>45</b>

# Table des figures

2.1	Architecture globale du système BI . . . . .	9
2.2	Modèle en étoile - Fait_Ventes et ses dimensions . . . . .	10
3.1	Saisonnalité implémentée (Ramadan +40%, Black Friday +150%) . . . . .	12
3.2	Distribution réelle de la segmentation RFM générée . . . . .	13
3.3	Logs d'exécution du script de génération des données . . . . .	15
3.4	Arborescence des fichiers générés par le script . . . . .	16
3.5	Exécution du script de chargement des données dans MySQL . . . . .	16
3.6	Liste des tables de la base <b>ecommerce_dw</b> dans phpMyAdmin . . . . .	17
4.1	Installation du driver MySQL ODBC ANSI . . . . .	19
4.2	Configuration du DSN MySQL . . . . .	19
4.3	Sélection de la source de données ODBC dans l'interface « Get Data » de Power BI. . . . .	20
4.4	Connexion ODBC depuis Power BI . . . . .	20
4.5	Sélection des tables relationnelles . . . . .	21
4.6	Sélection et conversion du noeud sessions du fichier JSON en table Power Query . . . . .	22
4.7	Expansion des enregistrements JSON en colonnes analytiques via Expand Columns . . . . .	22
4.8	Résultat de l'expansion des enregistrements JSON en table de faits de trafic web . . . . .	22
4.9	Suppression des doublons . . . . .	23
4.10	Application de Text.Proper() pour normaliser les noms clients . . . . .	23
4.11	Remplacement des valeurs "Ville inconnue" . . . . .	24
4.12	Conversion explicite en type Date pour la colonne DateInscription . . . . .	24
4.13	Résultat Table Date après nettoyage . . . . .	25
4.14	Configuration du Merge (Left Outer Join) . . . . .	25
4.16	Détection des ventes orphelines (ID_Client inexistant dans dim_client) . . . . .	26
4.15	Configuration du Merge (Left Outer Join) . . . . .	26
4.17	Fenêtre <i>Manage Relationships</i> dans Power BI . . . . .	28
5.1	Exemple d'affichage dynamique du produit leader . . . . .	33
5.2	Entonnoir de conversion e-commerce . . . . .	35
5.3	Vue d'ensemble des dossiers de mesures DAX implémentés dans le modèle Power BI. . . . .	37
6.1	Dashboard 1 - Vue Exécutive : KPI globaux et répartition par canal . . . . .	39

6.2	Dashboard 2 - Analyse Clients & Fidélité : Segmentation RFM et performance par ville . . . . .	40
6.3	Dashboard 3 - Performance Produits : Analyse ABC et contribution au CA	41
6.4	Dashboard 4 - Retours & Logistique : Analyse des retours et indice qualité	42
6.5	Dashboard 5 - Trafic Web & Conversion : Analyse du parcours client digital	43

# Liste des tableaux

1.1	Domaines analytiques couverts . . . . .	8
2.1	Description des tables de faits . . . . .	10
3.1	Volumes de données générées . . . . .	12
3.2	Formats de données générés et justification métier . . . . .	14
4.1	Stratégies de traitement des valeurs NULL . . . . .	24
5.1	Interprétation de la classification ABC . . . . .	34
5.2	Récapitulatif des mesures DAX implémentées et niveaux de complexité . . . . .	36
6.1	Structure des dashboards Power BI . . . . .	38

# Chapitre 1

## Introduction

### 1.1 Contexte du Projet

Dans un contexte de transformation digitale accélérée, les entreprises marocaines de e-commerce font face à des défis majeurs en matière d'analyse de données multi-canaux. Le projet **OmniMarket 360** répond à cette problématique en proposant une solution complète de Business Intelligence permettant d'analyser les performances commerciales à travers les canaux Web, Mobile et Magasins physiques.

Notre solution s'inspire des plateformes leaders du marché marocain (Jumia, Amazon Maroc) et intègre les spécificités locales telles que les périodes de soldes, le Ramadan, et le Black Friday qui influencent significativement les comportements d'achat.

### 1.2 Objectifs du Projet

Ce projet vise à démontrer notre maîtrise des concepts fondamentaux de la Business Intelligence à travers :

- La conception d'un **modèle dimensionnel en étoile** respectant les bonnes pratiques du domaine
- L'intégration de **sources de données hétérogènes** (SQL, Excel, CSV, JSON, XML)
- La mise en œuvre d'un processus **ETL complet** avec nettoyage et transformation des données
- La création de **dashboards interactifs** dans Power BI avec mesures DAX avancées
- L'analyse de **patterns métier** (saisonnalité, segmentation RFM, analyse de cohortes)

### 1.3 Périmètre Fonctionnel

Le système couvre les domaines analytiques suivants :

TABLE 1.1 – Domaines analytiques couverts

<b>Domaine</b>	<b>Indicateurs clés</b>
Ventes	CA par canal, panier moyen, taux de conversion
Produits	ABC Analysis, rotation des stocks, marges
Clients	Segmentation RFM, lifetime value, rétention
Logistique	Taux de retour, délais de livraison, transporteurs
Web Analytics	Trafic, taux d'abandon panier, funnel de conversion

# Chapitre 2

## Architecture Technique du Système

### 2.1 Architecture Globale

Notre architecture suit le paradigme classique des entrepôts de données avec une séparation claire entre les systèmes sources, la couche d'intégration ETL, et la couche de présentation.

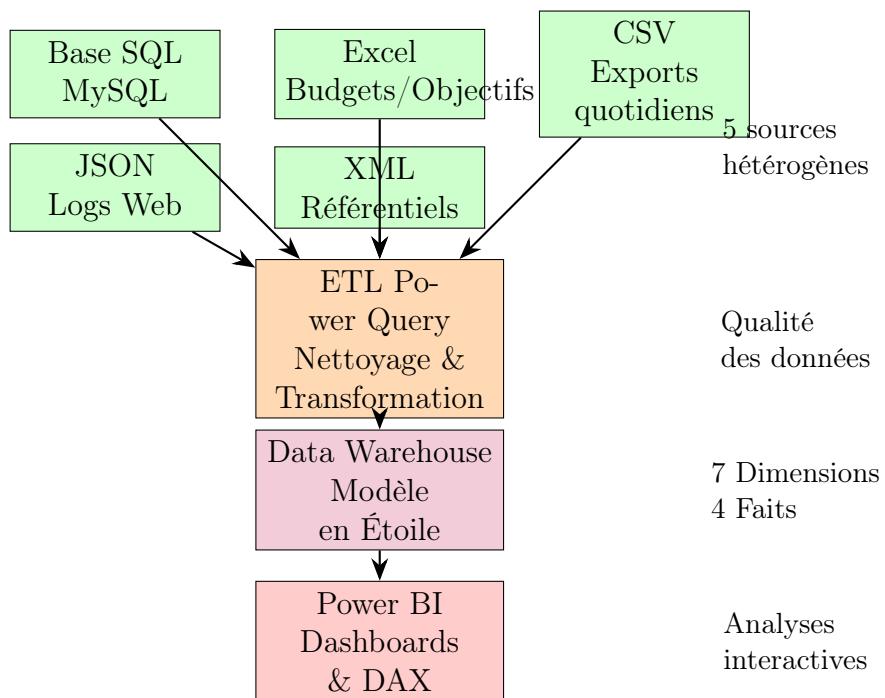


FIGURE 2.1 – Architecture globale du système BI

### 2.2 Modèle Dimensionnel

#### 2.2.1 Conception du Schéma en Étoile

Nous avons opté pour un modèle en étoile multi-faits permettant d'analyser différentes granularités business. Ce choix architectural offre :

- **Performance optimale** pour les requêtes analytiques
- **Simplicité de compréhension** pour les utilisateurs métier

— Flexibilité pour l'ajout de nouvelles dimensions

### 2.2.2 Tables de Faits

TABLE 2.1 – Description des tables de faits

Table	Description	Granularité
Fait_Ventes	Transactions commerciales tous canaux	Transaction
Fait_Retours	Retours produits avec motifs	Retour
Fait_Trafic_Web	Sessions utilisateurs site/app	Session
Fait_Stock	Snapshots hebdomadaires stocks	Produit/Semaine

### 2.2.3 Tables de Dimensions

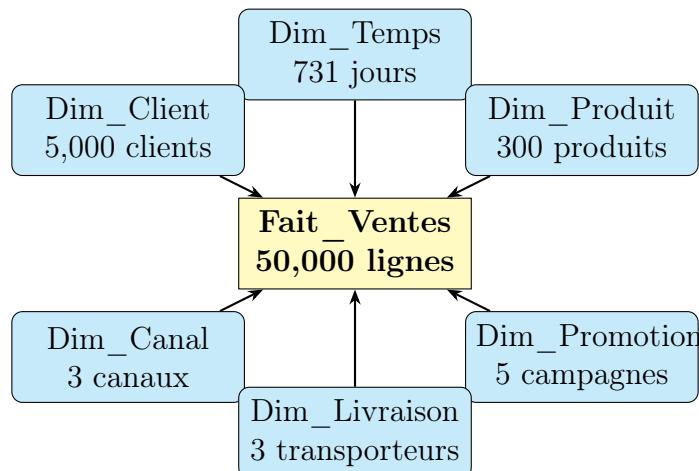


FIGURE 2.2 – Modèle en étoile - Fait\_Ventes et ses dimensions

### 2.2.4 Dimension Temporelle Enrichie

La dimension temps est cruciale pour les analyses de saisonnalité e-commerce. Nous l'avons enrichie avec :

```

1 Dim_Temps:
2   - ID_Date (PK)
3   - Date_Complete
4   - Annee, Trimestre, Mois, Semaine, Jour
5   - Jour_Semaine, Mois_Nom
6   - Est_Weekend (0/1)
7   - Est_Ferie (0/1)
8   - Saison_Commerciale (Normal, Soldes, Black Friday, Ramadan)

```

Listing 2.1 – Structure de Dim\_Temps

Cette granularité permet des analyses temporelles avancées comme l'identification des heures de pointe (crucial pour le e-commerce).

# Chapitre 3

## Génération et Intégration des Données

### 3.1 Introduction et Objectifs

La génération de données constitue une étape clé de ce projet, en l'absence de données réelles issues d'une plateforme e-commerce existante. L'objectif est de produire un jeu de données synthétiques **réalistes, cohérents et exploitables analytiquement**, tout en respectant les contraintes métier du commerce électronique.

Cette phase vise à :

- Simuler des comportements clients réalistes (saisonnalité, canaux, fidélité)
- Générer des volumes de données significatifs pour tester les performances analytiques
- Mettre en place une architecture multi-sources proche d'un contexte entreprise réel
- Préparer volontairement des anomalies afin d'évaluer les capacités de nettoyage ETL

### 3.2 Stratégie de Génération

Pour simuler un environnement e-commerce réaliste, nous avons développé un générateur de données en Python reposant sur des règles métier inspirées de pratiques observées dans des plateformes de vente en ligne réelles. L'objectif n'est pas de produire des données purement aléatoires, mais de générer des informations **statistiquement cohérentes, analytiquement exploitables et alignées avec les usages métiers**.

La génération repose sur une combinaison de distributions statistiques, de règles conditionnelles et de contraintes fonctionnelles permettant de reproduire des comportements clients crédibles (tels que la saisonnalité, la fidélité client et la diversité des canaux de vente).

#### 3.2.1 Volumes Générés

Les volumes générés ont été volontairement dimensionnés afin de se rapprocher d'un contexte d'entreprise de taille intermédiaire. Ce choix permet de tester à la fois :

- La capacité des outils décisionnels à traiter des volumes significatifs
  - Les performances des traitements ETL
  - La pertinence des analyses statistiques et agrégations temporelles
- Le volume de transactions constitue la table centrale du modèle, tandis que les autres entités viennent enrichir l'analyse selon différents axes (client, produit, temps, canal).

TABLE 3.1 – Volumes de données générées

Entité	Lignes	Période
Clients	5,000	2023-2024
Produits	300	6 catégories
Transactions	50,000	2 ans
Retours	2,476	5% des ventes
Sessions Web	100,000	Taux conv. 50%
Snapshots Stock	31,500	Hebdomadaire

### 3.2.2 Patterns Métier Implémentés

#### Saisonnalité des Ventes

Nous avons modélisé des variations saisonnières réalistes pour le marché marocain :

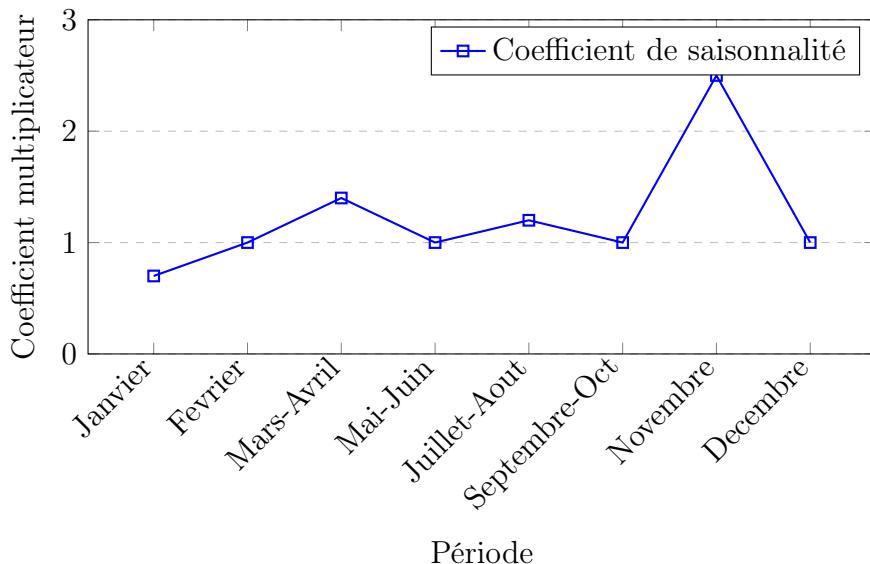


FIGURE 3.1 – Saisonnalité implémentée (Ramadan +40%, Black Friday +150%)

La saisonnalité implémentée permet de refléter les fluctuations naturelles de la demande observées dans le secteur du e-commerce marocain. Les périodes de forte activité, telles que le Black Friday ou le Ramadan, génèrent une augmentation significative du volume des ventes, tandis que certaines périodes creuses (début d'année) enregistrent une baisse notable.

Cette modélisation permet d'analyser l'impact des événements commerciaux sur le chiffre d'affaires, les volumes de ventes et le comportement d'achat des clients.

#### Distribution par Canal

Les ventes sont réparties selon les tendances actuelles du e-commerce :

- **Web** : 60% (panier moyen 450 MAD)
- **Mobile** : 30% (panier moyen 320 MAD, achats impulsifs)
- **Magasin** : 10% (panier moyen 680 MAD, achats réfléchis)

Cette répartition par canal reflète les tendances actuelles du commerce omnicanal. Le canal web reste dominant grâce à un parcours d'achat plus complet, tandis que le canal mobile est davantage associé à des achats rapides et impulsifs. Le canal magasin, bien que minoritaire en volume, présente un panier moyen plus élevé, souvent lié à des achats nécessitant conseil ou démonstration.

### 3.2.3 Segmentation RFM

La segmentation RFM (Recency, Frequency, Monetary) constitue un outil fondamental du marketing analytique. Elle permet de classer les clients selon leur valeur et leur niveau d'engagement, afin de faciliter des analyses telles que :

- L'identification des clients à forte valeur ajoutée
- Le ciblage de campagnes marketing personnalisées
- L'analyse de la fidélité et du churn potentiel

La segmentation Recency-Frequency-Monetary est calculée automatiquement :

```

1 def calculer_rfm(row):
2     if row['Nb_Achats'] >= 8 and row['CA_Total'] >= 5000:
3         return 'Gold' # 10% des clients
4     elif row['Nb_Achats'] >= 4:
5         return 'Silver' # 25% des clients
6     elif row['Nb_Achats'] >= 1:
7         return 'Bronze' # 40% des clients
8     else:
9         return 'Nouveau' # 25% des clients

```

Listing 3.1 – Logique de segmentation RFM

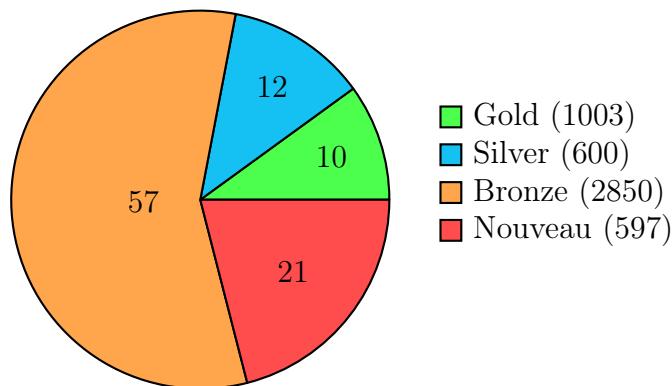


FIGURE 3.2 – Distribution réelle de la segmentation RFM générée

La distribution obtenue est volontairement déséquilibrée, conformément à la loi de Pareto, où une minorité de clients génère une part significative du chiffre d'affaires. Cette segmentation servira de base à plusieurs analyses décisionnelles dans les chapitres suivants.

### 3.2.4 Cohérence Statistique

Les distributions utilisées pour la génération des données ont été choisies afin de reproduire des comportements observés dans des plateformes e-commerce réelles :

- Distribution normale pour l'âge des clients, centrée autour de 35 ans
- Distribution de Poisson pour le nombre de pages vues par session

- Distribution exponentielle pour la durée des sessions web
  - Loi de Pareto (80/20) pour la concentration du chiffre d'affaires
- Ces choix permettent d'éviter des répartitions uniformes irréalistes et d'obtenir des données exploitables pour des analyses décisionnelles avancées.

### 3.3 Gestion des Anomalies et Cas Réels

Afin de simuler des contraintes rencontrées dans des systèmes réels, des anomalies volontaires ont été introduites lors de la génération des données. Ces anomalies permettent de tester la robustesse des processus ETL et des règles de nettoyage.

Les anomalies suivantes ont été intégrées volontairement :

- **50 doublons clients** (même email) pour tester la déduplication
- **200 valeurs NULL** sur champs non-critiques (téléphone, ville)
- **30 dates aberrantes** à corriger (ex : 2025-13-45)
- **Incohérences référentielles** (ID\_Client inexistant)

Ces anomalies sont utilisées comme cas de test pour la phase de nettoyage et de transformation des données dans le pipeline ETL.

### 3.4 Architecture Multi-Sources

Pour démontrer la maîtrise de l'intégration hétérogène, nous avons réparti les données sur 5 formats différents.

#### 3.4.1 Pourquoi Plusieurs Formats ?

Dans un contexte d'entreprise réel, les données proviennent rarement d'une source unique. Elles sont généralement issues de systèmes hétérogènes (ERP, CRM, fichiers Excel, API, bases transactionnelles). Afin de reproduire fidèlement cet environnement, nous avons volontairement généré les données sous plusieurs formats.

#### 3.4.2 Justification des Choix de Formats

TABLE 3.2 – Formats de données générés et justification métier

Format	Tables	Justification métier
Excel	Dim_Client, Fait_Ventes, Objectifs	Exports manuels des départements Vente et Marketing, budgets et reporting
CSV	Dim_Temps, Fait_Stock, Fait_Retours	Exports automatisés de systèmes legacy (ERP, WMS)
JSON	Dim_Canal, Dim_Livraison, Fait_Trafic_Web	Logs applicatifs et données issues d'outils analytics
XML	Dim_Produit, Référentiel_Géo, Dim_Motif_Retour	Catalogues fournisseurs, échanges B2B et flux EDI
SQL	Schéma complet	Base transactionnelle de production

## 3.5 Exécution du Script de Génération

### 3.5.1 Lancement et Logs d'Exécution

Le script principal de génération des données a été exécuté depuis le terminal via la commande suivante :

```
1 python generation_donnees.py
```

```

PROBLEMS ② OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\hp\Desktop\Projet_PowerBI_ECommerce\03_Scripts> python gen_data.py
❶ Démarrage génération des données E-Commerce (version multi-formats)...
❷ Génération Dim_Temps...
❸ 731 jours générés (2023-2024)

❹ Génération Dim_Client...
❺ 5050 clients générés (dont 50 doublons à nettoyer)
PS C:\Users\hp\Desktop\Projet_PowerBI_ECommerce\03_Scripts> python gen_data.py
❻ Démarrage génération des données E-Commerce (version multi-formats)...

❾ Génération Dim_Temps...
❿ 731 jours générés (2023-2024)

❻ Génération Dim_Client...
❽ 5050 clients générés (dont 50 doublons à nettoyer)
❾ Génération Dim_Temps...
❽ 731 jours générés (2023-2024)

❻ Génération Dim_Client...
❽ 5050 clients générés (dont 50 doublons à nettoyer)
❾ Génération Dim_Client...
❽ 5050 clients générés (dont 50 doublons à nettoyer)
❽ 5050 clients générés (dont 50 doublons à nettoyer)

❻ Génération Dim_Produit...
❽ 300 produits générés dans 6 catégories

❻ Génération Dim_Canal, Dim_Promotion, Dim_Livraison, Dim_Motif_Retour...
❽ Dimensions simples créées

❻ Génération Fait_Ventes...
❽ Cela peut prendre 1-2 minutes...
❻ 10000/50000 ventes générées...
❻ 20000/50000 ventes générées...
❻ 30000/50000 ventes générées...
❻ 40000/50000 ventes générées...
❻ 50000/50000 ventes générées...
❽ 50000 ventes générées

❻ Calcul segmentation RFM...
Distribution RFM: {'Bronze': 2850, 'Gold': 1003, 'Silver': 600, 'Nouveau': 597}

❻ Génération Fait_Retours...
❽ 2476 retours générés

❻ Génération Fait_Traffic_Web...
❻ 20000/100000 sessions générées...
❻ 40000/100000 sessions générées...
❻ 60000/100000 sessions générées...
❻ 80000/100000 sessions générées...
❻ 100000/100000 sessions générées...
❽ 100000 sessions web générées

❻ Génération Fait_Stock...
❽ 31500 enregistrements stock générés

❻ Export multi-sources...
❻ Export Excel : Dim_Client.xlsx + Fait_Ventes.xlsx + Objectifs_Mensuels.xlsx
❻ Export CSV : Dim_Temps, Dim_Promotion, Fait_Stock, Fait_Retours
❻ Export JSON : Dim_Canal, Dim_Livraison, Fait_Traffic_Web
❻ Export XML : Dim_Produit, Referentiel_Geo, Dim_Motif_Retour
❻ Export SQL : base_ventes.sql

❽ Export terminé !
❻ Dossier : ../02_Donnees/Sources/
❻ Excel : Dim_Client.xlsx, Fait_Ventes.xlsx, Objectifs_Mensuels.xlsx
❻ CSV : Dim_Temps.csv, Dim_Promotion.csv, Fait_Stock.csv, Fait_Retours.csv
❻ JSON : Dim_Canal.json, Dim_Livraison.json, Fait_Traffic_Web.json
❻ XML : Dim_Produit.xml, Referentiel_Geo.xml, Dim_Motif_Retour.xml
❻ SQL : base_ventes.sql

❽ Fin du script.

```

FIGURE 3.3 – Logs d'exécution du script de génération des données

### 3.5.2 Vérification de la Structure des Fichiers Générés

À l'issue de l'exécution, les fichiers générés sont automatiquement stockés dans le répertoire **02\_Donnees/Sources**. Une vérification visuelle a été réalisée afin de confirmer la présence et la cohérence de l'ensemble des fichiers produits.



FIGURE 3.4 – Arborescence des fichiers générés par le script

Cette structure confirme que l'export multi-sources a été réalisé avec succès, chaque format contenant les tables prévues.

## 3.6 Chargement des Données dans MySQL

### 3.6.1 Exécution du Script d'Upload SQL

Après la génération des fichiers, un second script Python (`upload_to_sql.py`) a été exécuté afin de créer le schéma de base de données et de charger les données dans MySQL.

```
1 python upload_to_sql.py
```

Le script affiche des messages de progression indiquant la création de la base, l'exécution du schéma SQL, le chargement des tables et la validation des volumes.

```
(env) PS C:\Users\hp\Desktop\Projet_PowerBI_ECommerce\03_Scripts> python upload_to_sql.py
↳ (1) Création base si nécessaire...
↳ (2) Connexion MySQL...
↳ (3) Exécution du schéma base_ventes.sql...
↳ (4) Lecture Excel : Dim_Client + Fait_Ventes...
↳ (5) Nettoyage/chargement : TRUNCATE puis INSERT...
C:\Users\hp\Desktop\Projet_PowerBI_ECommerce\03_Scripts\upload_to_sql.py:136: UserWarning: The provided table name 'Dim_Client' is not found exactly as such in the database after writing the table, possibly due to case sensitivity issues. Consider using lower case table names.
    df.to_sql()
C:\Users\hp\Desktop\Projet_PowerBI_ECommerce\03_Scripts\upload_to_sql.py:136: UserWarning: The provided table name 'Fait_Ventes' is not found exactly as such in the database after writing the table, possibly due to case sensitivity issues. Consider using lower case table names.
    df.to_sql()
↳ (6) Vérification volumes...
    - Dim_Client: 5050 lignes
    - Fait_Ventes: 50000 lignes
↳ Ventes avec ID_Client absent dans Dim_Client : 0 (attendu: 0)

Terminé : schéma créé + Dim_Client & Fait_Ventes chargées dans MySQL.
```

FIGURE 3.5 – Exécution du script de chargement des données dans MySQL

### 3.6.2 Validation dans phpMyAdmin

Une fois le chargement terminé, la base de données `ecommerce_dw` a été inspectée via l'outil phpMyAdmin afin de valider visuellement la création des tables et la cohérence des données.

The screenshot shows the phpMyAdmin interface with the following details:

- Left sidebar:** Shows the database tree with the `ecommerce_dw` database selected. It contains 11 tables: `dim_canal`, `dim_client`, `dim_livraison`, `dim_motif_retour`, `dim_produit`, `dim_promotion`, `dim_temps`, `fait_retours`, `fait_stock`, `fait_trafic_web`, and `fait_ventes`.
- Top navigation bar:** Includes tabs for Structure, SQL, Rechercher, Exporter, Importer, Opérations, Privileges, and Plus.
- Table list:** A table showing the 11 tables with their respective details:
 

Table	Action	Lignes	Type	Interclassement	Taille	Pt
dim_canal	Parcourir Structure Rechercher Insérer Vider Supprimer	5 059	InnoDB	utf8mb4_general_ci	1,5 Mio	utilisé
dim_client	Parcourir Structure Rechercher Insérer Vider Supprimer	5 059	InnoDB	utf8mb4_general_ci	16,0 kio	
dim_livraison	Parcourir Structure Rechercher Insérer Vider Supprimer	0	InnoDB	utf8mb4_general_ci	16,0 kio	
dim_motif_retour	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
dim_produit	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
dim_promotion	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
dim_temps	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
fait_retours	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
fait_stock	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
fait_trafic_web	Parcourir Structure Rechercher Insérer Vider Supprimer	8	InnoDB	utf8mb4_general_ci	16,0 kio	
fait_ventes	Parcourir Structure Rechercher Insérer Vider Supprimer	3,5 Mio	InnoDB	utf8mb4_general_ci	5,2 Mio	

FIGURE 3.6 – Liste des tables de la base `ecommerce_dw` dans phpMyAdmin

## Conclusion du Chapitre

Ce chapitre a présenté le processus complet de génération et d'intégration des données synthétiques utilisées dans ce projet. À travers une architecture multi-sources, des volumes réalistes et des mécanismes avancés de simulation métier, nous avons constitué une base de données cohérente et exploitable.

Le chapitre suivant sera consacré à la mise en place du processus ETL, incluant les étapes de nettoyage, transformation et chargement des données vers l'entrepôt décisionnel.

# Chapitre 4

## Processus ETL et Intégration

### 4.1 Introduction

Ce chapitre présente de manière unifiée et cohérente l'ensemble du processus technique mis en œuvre dans le cadre de ce projet décisionnel. Il couvre successivement l'architecture technique, la connectivité aux sources de données, le processus ETL (Extraction, Transformation, Chargement) ainsi que la modélisation finale sous Power BI.

L'objectif est de démontrer une maîtrise complète de la chaîne BI, allant de l'acquisition multi-sources à la structuration d'un modèle analytique robuste, performant et conforme aux bonnes pratiques du décisionnel.

### 4.2 Architecture Technique et Connectivité

#### 4.2.1 Environnement technique

La solution BI repose sur une architecture locale structurée autour des composants suivants :

- Un SGBD relationnel **MySQL** hébergeant la base décisionnelle `ecommerce_dwh`
  - L'outil de reporting et de modélisation Power BI Desktop (64 bits)
  - Le moteur ETL intégré Power Query
  - Une connectivité standardisée via ODBC

Cette architecture permet une séparation claire entre la couche stockage (MySQL), la couche transformation (Power Query) et la couche analytique (Modèle Power BI).

#### 4.2.2 Choix et configuration de la connectivité ODBC

Le standard **ODBC** a été retenu afin de garantir :

- Une compatibilité native avec Power BI
- Une abstraction du SGBD sous-jacent
- Une stabilité accrue des échanges de données

Le driver **MySQL Connector/ODBC 9.5 ANSI** a été installé. Le choix de la version ANSI est motivé par sa meilleure gestion des types numériques et l'absence de conflits d'encodage dans Power BI Desktop.

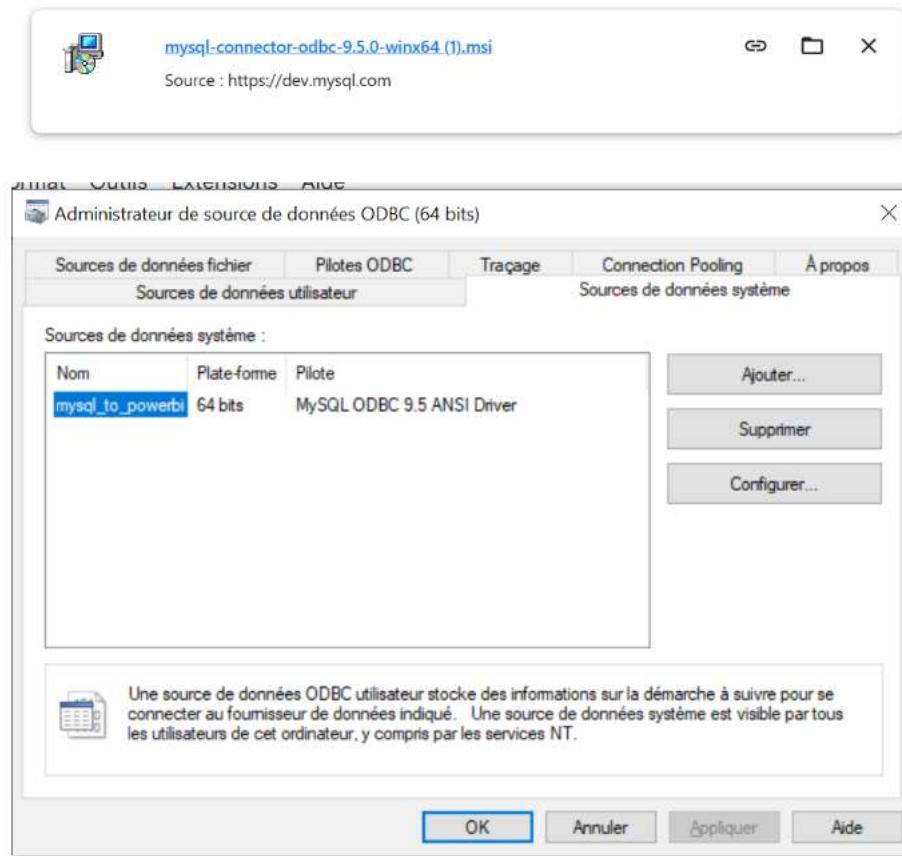


FIGURE 4.1 – Installation du driver MySQL ODBC ANSI

#### 4.2.3 Crédation du DSN et connexion à Power BI

Un **DSN système** a été configuré afin de centraliser les paramètres de connexion à la base MySQL.

- Serveur : localhost
- Base : `ecommerce_dw`
- Port : 3306
- Authentification MySQL sécurisée

Cette approche garantit une maintenance simplifiée et une réutilisation facilitée de la connexion.



FIGURE 4.2 – Configuration du DSN MySQL

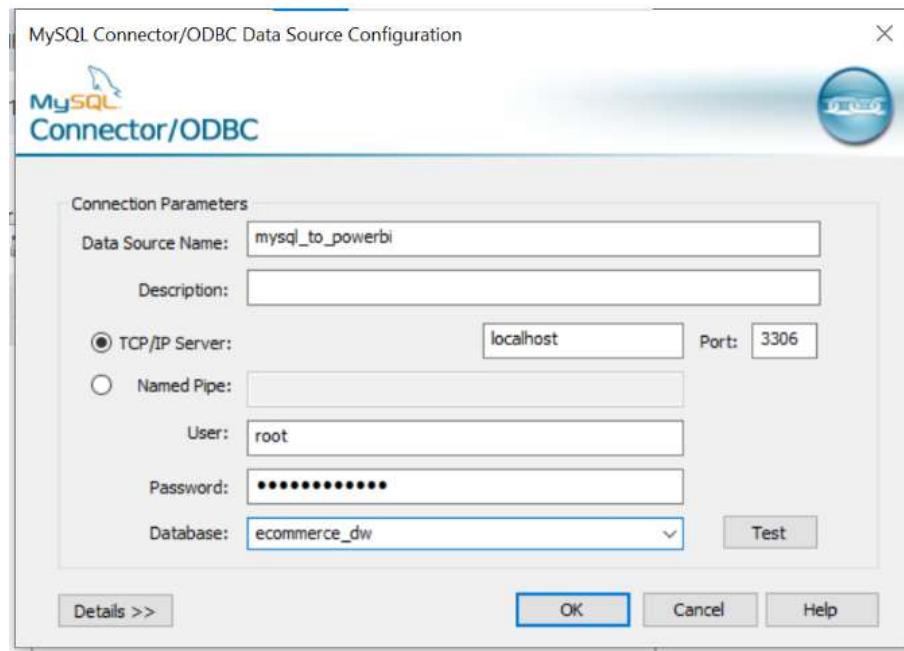


FIGURE 4.3 – Sélection de la source de données ODBC dans l’interface « Get Data » de Power BI.

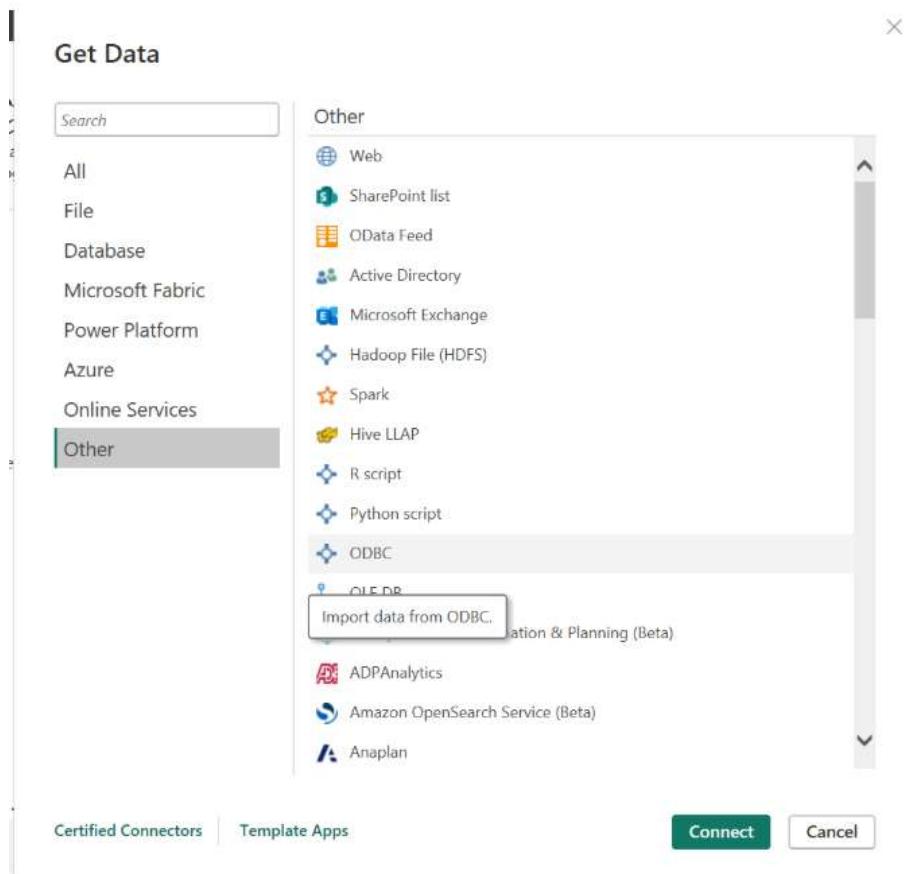


FIGURE 4.4 – Connexion ODBC depuis Power BI

## 4.3 Extraction et Acquisition Multi-sources

### 4.3.1 Extraction relationnelle (SQL / MySQL)

Les tables de faits et de dimensions sont extraites depuis MySQL en **mode Import**, garantissant de meilleures performances analytiques.

```

1 let
2 Source = Sql.Database("localhost", "ecommerce_dw"),
3 Fait_Ventes = Source{[Item="Fait_Ventes"]}[Data]
4 in
5 Fait_Ventes

```

Listing 4.1 – Connexion SQL dans Power Query (M)

Cette extraction est volontairement sélective afin de limiter le volume chargé et de respecter les principes de performance BI.

ID_Client	Code_Client	Prenom	Nom	Nom_Complet
1	CU00001	Khadija	Alami	Khadija Alami
2	CU00002	Nadia	Benjelloun	Nadia Benjelloun
3	CU00003	Bilal	Mekouar	Bilal Mekouar
4	CU00004	Sara	Tazi	Sara Tazi
5	CU00005	Youssef	Tounsi	Youssef Tounsi
6	CU00006	Karim	Benjelloun	Karim Benjelloun
7	CU00007	Omar	Yacoubi	Omar Yacoubi
8	CU00008	Mehdi	Benjelloun	Mehdi Benjelloun
9	CU00009	Fatima	Alami	Fatima Alami
10	CU00010	Khadija	Idrissi	Khadija Idrissi
11	CU00011	Aicha	Sefrioui	Aicha Sefrioui
12	CU00012	Nadia	Tounsi	Nadia Tounsi
13	CU00013	Nadia	Tazi	Nadia Tazi
14	CU00014	Karim	Idrissi	Karim Idrissi
15	CU00015	Tazi	Karim	Karim Tazi
16	CU00016	Zineb	Tounsi	Zineb Tounsi
17	CU00017	Zineb	El Fassi	Zineb El Fassi
18	CU00018	Youssef	Yacoubi	Youssef Yacoubi
19	CU00019	Rachid	Lahliou	Rachid Lahliou
20	CU00020	Nadia	Yacoubi	Nadia Yacoubi
21	CU00021	Youssef	Kettani	Youssef Kettani
22	CU00022	Bilal	Benjelloun	Bilal Benjelloun

FIGURE 4.5 – Sélection des tables relationnelles

### 4.3.2 Intégration des données JSON (Trafic Web)

Les données de trafic web sont fournies sous forme semi-structurée (JSON). Power Query est utilisé pour aplatisir ces structures imbriquées.

```

1 let
2 Source = Json.Document(File.Contents("Fait_Trafic_Web.json")),
3 Sessions = Source[sessions],
4 ToTable = Table.FromList(Sessions, Splitter.SplitByNothing())
5 in
6 ToTable

```

Listing 4.2 – Parsing JSON dans Power Query

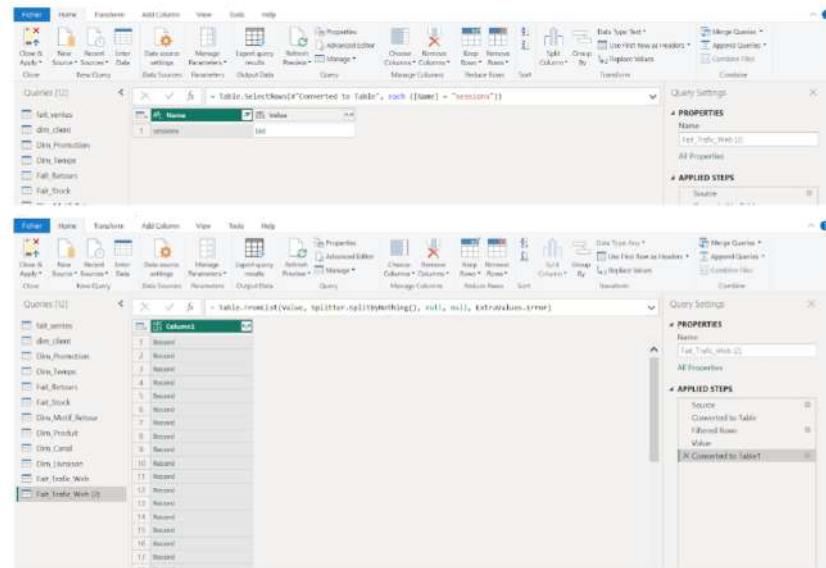


FIGURE 4.6 – Sélection et conversion du nœud sessions du fichier JSON en table Power Query

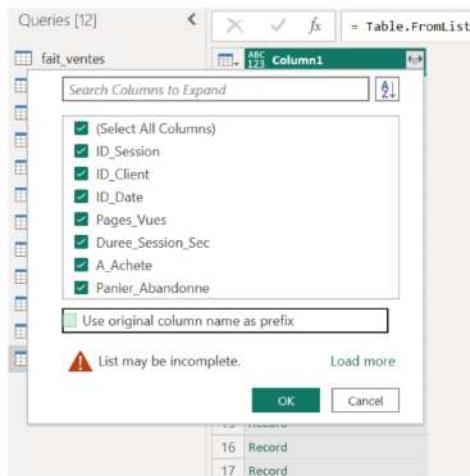


FIGURE 4.7 – Expansion des enregistrements JSON en colonnes analytiques via Expand Columns

ID_Session	ID_Client	ID_Date	Pages_Vues	Duree_Session_Sec
1	1	null	422	8
2	2	null	211	3
3	3	null	669	4
4	4	1494	212	3
5	5	2810	615	4
6	6	2057	349	2
7	7	null	112	4
8	8	465	310	4
9	9	null	536	2
10	10	null	290	3
11	11	null	345	4
12	12	4606	69	5
13	13	1933	283	7
14	14	null	332	4

FIGURE 4.8 – Résultat de l'expansion des enregistrements JSON en table de faits de trafic web

### 4.3.3 Intégration des données XML

Les dimensions Produits et Géographie sont fournies au format XML.

```

1 let
2 Source = Xml.Tables(File.Contents("Dim_Produit.xml")),
3 Produit = Source{0}[Produit],
4 Expanded = Table.ExpandTableColumn(Produit, "Element:Text")
5 in
6 Expanded

```

Listing 4.3 – Extraction XML des produits

## 4.4 Transformation et Qualité des Données

### 4.4.1 Nettoyage et standardisation

Une phase de nettoyage systématique est appliquée :

- Suppression des doublons clients
- Normalisation des champs texte
- Typage explicite des colonnes

The screenshot shows the Power BI Advanced Editor interface. A table with columns 'Email', 'Telephone', and 'Date\_Inscription' is displayed. The 'Remove Rows' option is selected from the context menu, which lists various row cleanup actions.

Email	Telephone	Date_Inscription
khadija.alami251@email.ma	212628728463	21/01/2023 Tanger
nadia.benjelloun33@email.ma	212622575562	02/01/2023 Tanger
bilal.mekouar575@email.ma	212697226012	

FIGURE 4.9 – Suppression des doublons

The screenshot shows the Power BI Advanced Editor interface. A table with columns 'Email' and 'Telephone' is displayed. The 'Format' option is selected from the context menu, which lists text transformation functions such as 'lowercase', 'uppercase', 'Capitalize Each Word', 'Trim', 'Clean', 'Add Prefix', and 'Add Suffix'.

Email	Telephone
khadija.alami251@email.ma	212628728463
nadia.benjelloun33@email.ma	212622575562
bilal.mekouar575@email.ma	212697226012
sara.tazi604@email.ma	212710877248

FIGURE 4.10 – Application de Text.Proper() pour normaliser les noms clients

#### 4.4.2 Gestion des valeurs manquantes

TABLE 4.1 – Stratégies de traitement des valeurs NULL

Colonne	Stratégie	Justification
"Non renseigné"	Champ optionnel	Telephone
Maintien cohérence géographique ID <sub>Livraison</sub>	NULL conservé	Imputation / Replace Value
Suppression ligne	Donnée critique	Cas ventes magasin Montage

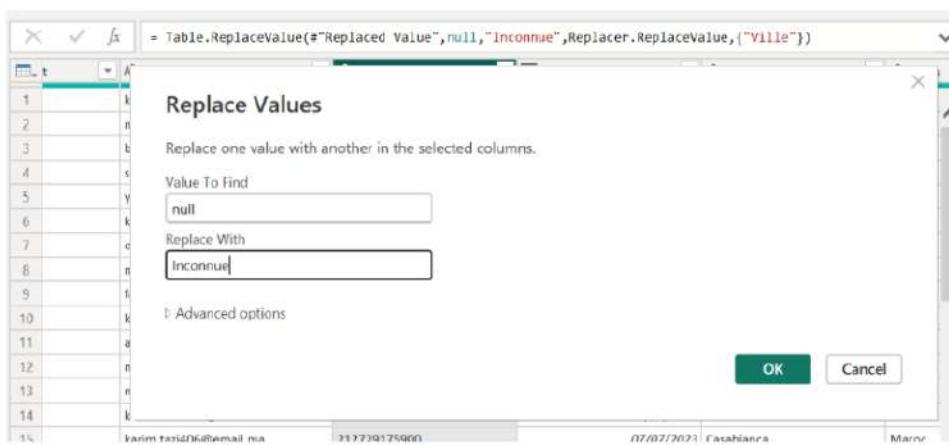
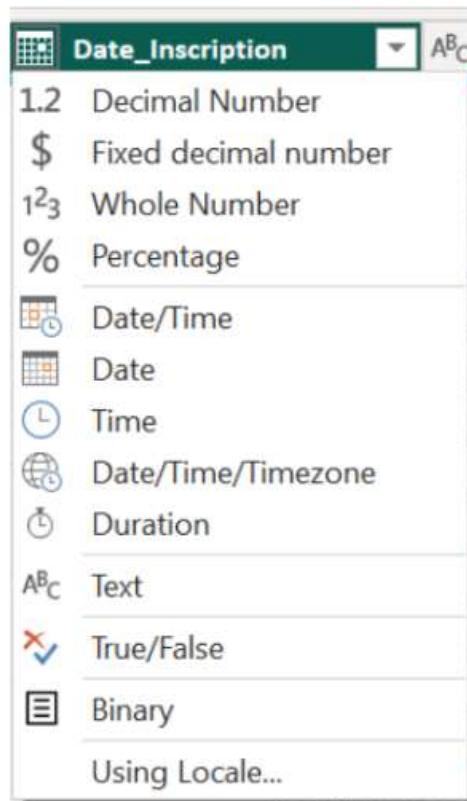


FIGURE 4.11 – Remplacement des valeurs "Ville inconnue"

FIGURE 4.12 – Conversion explicite en type Date pour la colonne Date<sub>Inscription</sub>

Date_Complete	Année	Trimestre	Mois
01/01/2023	2023	1	
02/01/2023	2023	1	
03/01/2023	2023	1	
04/01/2023	2023	1	
05/01/2023	2023	1	
06/01/2023	2023	1	
07/01/2023	2023	1	
08/01/2023	2023	1	
09/01/2023	2023	1	
10/01/2023	2023	1	
11/01/2023	2023	1	
12/01/2023	2023	1	
13/01/2023	2023	1	

FIGURE 4.13 – Résultat Table Date après nettoyage

#### 4.4.3 Contrôle de l'intégrité référentielle

Des jointures de type *Left Anti Join* ont été utilisées afin d'identifier et supprimer les lignes orphelines.

Merge

Select a table and matching columns to create a merged table.

fait\_ventes

ID_Vente	ID_Client	ID_Produit	ID_Date	Date_Vente	Heure_Vente	DateTime_Vente	ID_Canal
1	3815	247	214	02/08/2023	16	02/08/2023 16:12:43	1
2	99	59	607	29/08/2024	18	29/08/2024 18:19:02	1
3	580	265	199	18/07/2023	10	18/07/2023 10:13:31	1
4	393	78	700	30/11/2024	0	30/11/2024 00:02:51	2

dim\_client

ID_Client	Code_Client	Prenom	Nom	Nom_Complet	Email	Telephone	Dat
1	CLI00001	Khadija	Alami	Khadija Alami	khadija.alami251@email.ma	212628728463	
2	CLI00002	Nadia	Benjelloun	Nadia Benjelloun	nadia.benjelloun33@email.ma	212622575562	
3	CLI00003	Bilal	Mekouar	Bilal Mekouar	bilal.mekouar575@email.ma	212697226012	
4	CLI00004	Sara	Tazi	Sara Tazi	sara.tazi604@email.ma	212710872248	

Join Kind

Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

Fuzzy matching options

The selection matches 50000 of 50000 rows from the first table.

OK Cancel

FIGURE 4.14 – Configuration du Merge (Left Outer Join)

ID_Client	Code_Client	Prenom	Nom
1	CLI00001	Khadija	Alami
6	99 CLI00099	Laila	Lahbou
3	99 CLI00099	Laila	Lahbou
8	580 CLI00580	Rachid	El Fassi
9	3 CLI00003	Bilal	Mekouar
3	393 CLI00393	Fatima	El Fassi
6	4 CLI00004	Sara	Tazi
1	4 CLI00004	Sara	Tazi
8	4 CLI00004	Sara	Tazi
2	5 CLI00005	Youssef	Tounsi
0	6 CLI00006	Karim	Benjelloun
6	767 CLI00767	Samira	Tazi
1	767 CLI00767	Samira	Tazi

FIGURE 4.16 – Détection des ventes orphelines (ID\_Client inexistant dans dim\_client)

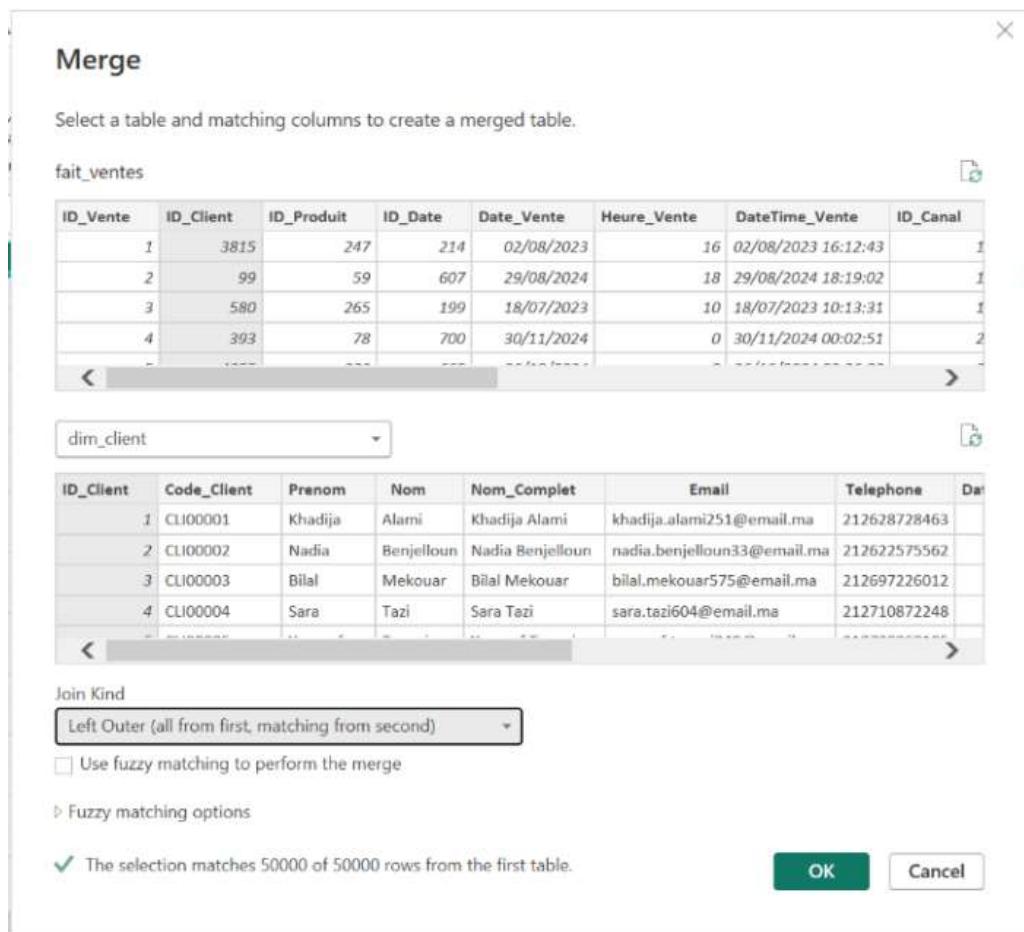


FIGURE 4.15 – Configuration du Merge (Left Outer Join)

## 4.5 Enrichissement et Calculs ETL

### 4.5.1 Colonnes dérivées

```

1 = Table.AddColumn(Source, "Marge_Pct",
2 each if [Montant_HT] <> 0
3 then ([Marge] / [Montant_HT]) * 100
4 else 0)

```

Listing 4.4 – Ajout d'une colonne MargePct

Ces enrichissements permettent de déplacer une partie du calcul en amont, réduisant la charge DAX.

## 4.6 Modélisation et Schéma Relationnel

### 4.6.1 Schéma en étoile

Le modèle décisionnel repose sur un **schéma en étoile**, centré autour des tables de faits et de dimensions.

### 4.6.2 Gestion des relations

- Cardinalité 1 :N
- Filtrage unidirectionnel
- Absence de relations ambiguës

**Manage relationships**

The screenshot shows the 'Manage relationships' window in Power BI. At the top, there are buttons for '+ New relationship', 'Autodetect', 'Edit', 'Delete', and 'Filter'. The main area is a table with columns: 'From: table (column)', 'Relationship', 'To: table (column)', and 'Status'. There are 18 rows, each representing a relationship. Most rows have a status of 'Active'. Some rows are highlighted with a grey background, such as the second row ('Fait\_Retours (ID\_Motif)') and the fifth row ('Fait\_Stock (ID\_Produit)'). The 'Relationship' column shows cardinalities: \*->□—1 for most rows, except for the third row which shows ∑->□—1.

From: table (column)	Relationship	To: table (column)	Status
Fait_Retours (ID_Date_Retour)	*->□—1	Dim_Temps (ID_Date)	Active
Fait_Retours (ID_Motif)	*->□—1	Dim_Motif_Retour (ID_Motif)	Active
∑ Fait_Retours (ID_Vente)	*->□—1	∑ fait_ventes (ID_Vente)	Inactive
Fait_Stock (ID_Date)	*->□—1	Dim_Temps (ID_Date)	Active
Fait_Stock (ID_Produit)	*->□—1	Dim_Produit (ID_Produit)	Active
fait_ventes (ID_Canal)	*->□—1	Dim_Canal (ID_Canal)	Active
fait_ventes (ID_Client)	*->□—1	dim_client (ID_Client)	Active
fait_ventes (ID_Date)	*->□—1	Dim_Temps (ID_Date)	Active
fait_ventes (ID_Livraison)	*->□—1	Dim_Livraison (ID_Livraison)	Active
fait_ventes (ID_Produit)	*->□—1	Dim_Produit (ID_Produit)	Active
fait_ventes (ID_Canal)	*->□—1	Dim_Canal (ID_Canal)	Active
fait_ventes (ID_Client)	*->□—1	dim_client (ID_Client)	Active
fait_ventes (ID_Date)	*->□—1	Dim_Temps (ID_Date)	Active
fait_ventes (ID_Livraison)	*->□—1	Dim_Livraison (ID_Livraison)	Active
fait_ventes (ID_Produit)	*->□—1	Dim_Produit (ID_Produit)	Active
fait_ventes (ID_Promotion)	*->□—1	Dim_Promotion (ID_Promotion)	Active

**Close**

FIGURE 4.17 – Fenêtre *Manage Relationships* dans Power BI

## Conclusion du Chapitre

Ce chapitre a démontré la maîtrise complète du processus BI, allant de l'architecture technique à la modélisation analytique. Le processus ETL mis en œuvre garantit des données fiables, cohérentes et optimisées pour l'analyse décisionnelle.

# Chapitre 5

## Mesures DAX et Intelligence Analytique

### 5.1 Architecture des Mesures

L'ensemble des mesures DAX a été structuré en \*\*6 catégories fonctionnelles\*\* pour assurer maintenabilité et cohérence analytique :

1. **Ventes de base** : CA, quantités, transactions
2. **Comparaisons temporelles** : YoY, YTD, moyennes mobiles
3. **Clients** : rétention, valeur par client
4. **Produits** : ranking, contribution, analyse ABC
5. **Retours** : taux de retour, impact financier
6. **Trafic web** : conversion, abandon panier

### 5.2 Mesures de Base

#### 5.2.1 Indicateurs de Ventes

Les mesures fondamentales permettent de suivre le chiffre d'affaires et les volumes :

```
1 CA_Total =
2 SUM(fait_ventes[Montant_TTC])
3
4 CA_HT =
5 SUM(fait_ventes[Montant_HT])
6
7 Nb_Transactions =
8 COUNTROWS(fait_ventes)
9
10 Panier_Moyen =
11 DIVIDE([CA_Total], [Nb_Transactions], 0)
12
13 Quantite_Vendue =
14 SUM(fait_ventes[Quantite])
```

Listing 5.1 – Mesures DAX - Ventes de base

## 5.2.2 Indicateurs de Rentabilité

Le suivi des marges et remises permet d'évaluer la profitabilité réelle :

```

1 Marge_Totale =
2 SUM(fait_ventes[Marge])
3
4 Taux_Marge =
5 DIVIDE([Marge_Totale], [CA_Total], 0)
6
7 Marge_Moyenne =
8 AVERAGE(fait_ventes[Marge])
9
10 Remise_Totale =
11 SUM(fait_ventes[Remise_Appliquee])
12
13 Taux_Remise_Moyen =
14 DIVIDE([Remise_Totale], [CA_HT], 0)
```

Listing 5.2 – Mesures DAX - Marges et Remises

## 5.3 Time Intelligence : Analyses Temporelles

### 5.3.1 Comparaisons Année sur Année (YoY)

Les mesures de comparaison permettent d'évaluer la croissance :

```

1 CA_Annee_Precedente =
2 CALCULATE(
3     [CA_Total],
4     SAMEPERIODLASTYEAR(dim_temps[Date_Complete])
5 )
6
7 Evolution_CA_YoY =
8 DIVIDE(
9     [CA_Total] - [CA_Annee_Precedente],
10    [CA_Annee_Precedente],
11    0
12 )
13
14 Evolution_CA_YoY_Pct =
15 FORMAT([Evolution_CA_YoY], "0.0%")
```

Listing 5.3 – Evolution Year-over-Year

### 5.3.2 Comparaisons Mois sur Mois

```

1 CA_Mois_Precedent =
2 CALCULATE(
3     [CA_Total],
4     PREVIOUSMONTH(dim_temps[Date_Complete])
5 )
6
7 Evolution_vs_Mois_Precedent =
8 DIVIDE(
9     [CA_Total] - [CA_Mois_Precedent],
```

```

10     [CA_Mois_Precedent] ,
11     0
12 )

```

Listing 5.4 – Evolution mensuelle

### 5.3.3 Cumuls Année en Cours (YTD)

Les mesures *Year-To-Date* permettent de suivre la progression annuelle :

```

1 CA_YTD =
2 TOTALYTD(
3     [CA_Total],
4     dim_temps[Date_Complete]
5 )
6
7 CA_YTD_Annee_Precedente =
8 CALCULATE(
9     [CA_YTD],
10    SAMEPERIODLASTYEAR(dim_temps[Date_Complete])
11 )

```

Listing 5.5 – Chiffre d'affaires cumulé YTD

### 5.3.4 Moyennes Mobiles

La moyenne mobile sur 3 mois lisse les variations saisonnières :

```

1 CA_MA_3_Mois =
2 CALCULATE(
3     [CA_Total],
4     DATESSINPERIOD(
5         dim_temps[Date_Complete],
6         LASTDATE(dim_temps[Date_Complete]),
7         -3,
8         MONTH
9     )
10 ) / 3

```

Listing 5.6 – Moving Average 3 mois

#### Note méthodologique

Cette mesure est essentielle pour détecter les **tendances sous-jacentes** en éliminant le bruit des fluctuations mensuelles.

## 5.4 Analyses Clients

### 5.4.1 Indicateurs de Base

```

1 Nb_Clients_Actifs =
2 DISTINCTCOUNT(fait_ventes[ID_Client])
3
4 CA_Par_Client =

```

```
5 DIVIDE([CA_Total], [Nb_Clients_Actifs], 0)
```

Listing 5.7 – Métriques clients

## 5.4.2 Taux de Rétention

Le taux de rétention mesure la fidélité client d'une année sur l'autre :

```
1 Taux_Retention =
2 VAR ClientsAnneeActuelle =
3     CALCULATE(
4         DISTINCTCOUNT(fait_ventes[ID_Client]),
5         dim_temps[Annee] = YEAR(TODAY())
6     )
7 VAR ClientsAnneePrecedente =
8     CALCULATE(
9         DISTINCTCOUNT(fait_ventes[ID_Client]),
10        dim_temps[Annee] = YEAR(TODAY()) - 1
11    )
12 RETURN
13 DIVIDE(ClientsAnneeActuelle, ClientsAnneePrecedente, 0)
```

Listing 5.8 – Rétention annuelle

### Limitation DAX

Cette mesure utilise **ALL()** et n'est **pas compatible avec les slicers temporels**. Elle doit être utilisée dans un visuel isolé sans filtres de date.

## 5.5 Analyse Produits : Ranking et ABC

### 5.5.1 Classement des Produits

Le ranking permet d'identifier les best-sellers :

```
1 Rank_Produit_CA =
2 RANKX(
3     ALL(dim_produit[Nom_Produit]),
4     [CA_Total],
5     ,
6     DESC,
7     DENSE
8 )
```

Listing 5.9 – Ranking produits par CA

### 5.5.2 Identification du Produit Leader

La fonction **FIRSTNONBLANK()** permet d'extraire automatiquement le nom du produit générant le plus de CA :

```
1 Top_Produit_Nom =
2 FIRSTNONBLANK(
3     TOPN(1, VALUES(dim_produit[Nom_Produit]), [CA_Total], DESC),
4     1
```

5 )

### Listing 5.10 – Top produit dynamique

**Utilisation :** Cette mesure peut être affichée dans une carte KPI pour mettre en avant le best-seller du mois/trimestre sélectionné.

```
Top_Produit_Nom = "Ordinateur Portable XPS 15"
```

FIGURE 5.1 – Exemple d'affichage dynamique du produit leader

### 5.5.3 Contribution au Chiffre d’Affaires

```
1 Contribution_CA =
2 DIVIDE(
3     [CA_Total],
4     CALCULATE([CA_Total], ALL(dim_produit)),
5     0
6 )
```

### Listing 5.11 – Part de marché produit

### 5.5.4 Nombre de Produits Vendus

Cette mesure permet de suivre la diversité du catalogue effectivement commercialisé :

```
1 Nb_Produits_Vendus =
2 DISTINCTCOUNT(fait_ventes[ID_Produit])
```

### Listing 5.12 – Comptage distinct des produits vendus

### 5.5.5 Chiffre d’Affaires Moyen par Produit

Indicateur de performance unitaire des produits :

```
1 CA_Moyen_Produit =
2 DIVIDE([CA_Total], [Nb_Produits_Vendus], 0)
```

### Listing 5.13 – CA moyen par référence

#### Insight Métier

Un CA\_Moyen\_Produit élevé peut indiquer :

- Une stratégie de positionnement premium
- Une concentration des ventes sur peu de références
- Un besoin d’élargir le catalogue actif

### 5.5.6 Classification ABC (Loi de Pareto)

L’analyse ABC applique la règle 80/20 pour prioriser les actions :

```

1 Contribution_Cumul =
2 CALCULATE(
3     [Contribution_CA],
4     FILTER(
5         ALLSELECTED(dim_produit),
6         [Rank_Produit_CA] <= EARLIER([Rank_Produit_CA])
7     )
8 )
9
10 Classe_ABC =
11 SWITCH(
12     TRUE(),
13     [Contribution_Cumul] <= 0.7, "A - Top 70%",
14     [Contribution_Cumul] <= 0.9, "B - 70-90%",
15     "C - Bottom 10%"
16 )

```

Listing 5.14 – Analyse ABC avec contribution cumulée

TABLE 5.1 – Interprétation de la classification ABC

Classe	Contribution	Action Stratégique
A	0-70%	Produits prioritaires : maximiser la disponibilité
B	70-90%	Suivi régulier, optimisation logistique
C	90-100%	Évaluer pertinence du catalogue

## 5.6 Gestion des Retours

Les retours impactent directement la rentabilité et nécessitent un suivi rigoureux :

```

1 Nb_Retours =
2 COUNTROWS(fait_retours)
3
4 Taux_Retour =
5 DIVIDE([Nb_Retours], [Nb_Transactions], 0)
6
7 Montant_Retours =
8 SUM(fait_retours[Montant_Rembourse])
9
10 CA_Net_Retours =
11 [CA_Total] - [Montant_Retours]

```

Listing 5.15 – Mesures DAX - Retours

### Insight Métier

Un taux de retour > 5% signale généralement un problème qualité produit ou une inadéquation entre description et réalité.

## 5.7 Trafic Web et Conversion

### 5.7.1 Indicateurs de Performance Web

```

1 Nb_Sessions =
2 COUNTROWS(fait_trafic_web)
3
4 Nb_Sessions_Achat =
5 CALCULATE(
6     COUNTROWS(fait_trafic_web),
7     fait_trafic_web[A_Achete] = 1
8 )
9
10 Taux_Conversion =
11 DIVIDE([Nb_Sessions_Achat], [Nb_Sessions], 0)

```

Listing 5.16 – Sessions et achats

### 5.7.2 Abandon de Panier

L'abandon de panier représente une perte de revenu potentielle majeure :

```

1 Nb_Paniers_Abandonnees =
2 CALCULATE(
3     COUNTROWS(fait_trafic_web),
4     fait_trafic_web[Panier_Abandonne] = 1
5 )
6
7 Taux_Abandon_Panier =
8 DIVIDE([Nb_Paniers_Abandonnees], [Nb_Sessions], 0)
9
10 CA_Potentiel_Perdu =
11 [Nb_Paniers_Abandonnees] * [Panier_Moyen]

```

Listing 5.17 – Analyse abandon panier

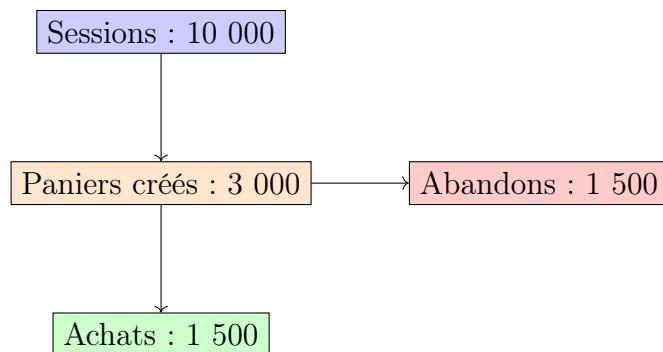


FIGURE 5.2 – Entonnoir de conversion e-commerce

## 5.8 Limitations et Bonnes Pratiques

### 5.8.1 Mesures Incompatibles avec les Filtres

Certaines mesures utilisent ALL(), EARLIER(), ou RANKX() sur des tables complètes, ce qui les rend **insensibles aux slicers** :

- Taux\_Retention (comparaison annuelle fixe)
- Rank\_Produit\_CA (ranking global)
- Classe\_ABC (cumul nécessitant contexte complet)

**Solution :** Utiliser ces mesures dans des visuels dédiés sans slicers temporels, ou créer des variantes avec ALLSELECTED().

### 5.8.2 Optimisation des Performances

1. **Éviter les calculs en ligne** : Créer des colonnes calculées en Power Query pour les transformations simples
2. **Utiliser Variables DAX** : Réduire les évaluations multiples
3. **Privilégier SUM() sur SUMX()** : Sauf si itération nécessaire

```

1 CA_Optimise =
2 VAR TotalVentes = [CA_Total]
3 VAR TotalAnnePrecedente = [CA_Annee_Precédente]
4 RETURN
5 DIVIDE(TotalVentes - TotalAnnePrecedente, TotalAnnePrecedente, 0)

```

Listing 5.18 – Exemple d’optimisation avec variables

## 5.9 Synthèse des Mesures Clés

TABLE 5.2 – Récapitulatif des mesures DAX implémentées et niveaux de complexité

Catégorie	Mesure	Complexité
Ventes	CA_Total, Panier_Moyen	★○○
Temporelles	CA_YTD, Evolution_CA_YoY	★★○
Clients	Taux_Retention [cite : 1042]	★★★
Produits (base)	Quantite_Vendue, Marge_Moyenne	★○○
Produits (avancé)	Classe_ABC, Rank_Produit_CA	★★★
Retours	CA_Net_Retours, Taux_Retour	★○○
Web	Taux_Conversion, CA_Potentiel_Perdu	★★○

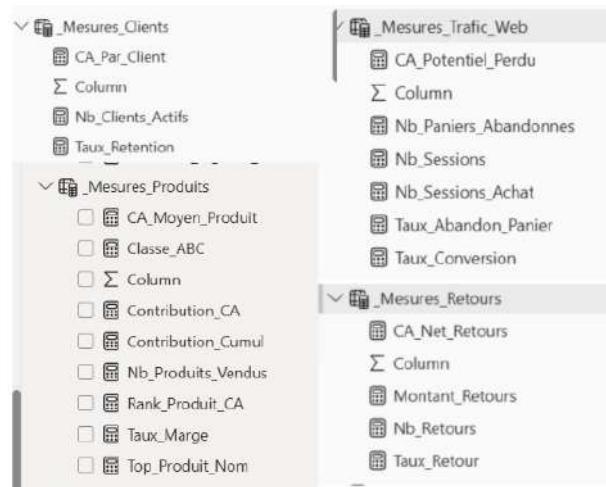


FIGURE 5.3 – Vue d’ensemble des dossiers de mesures DAX implémentés dans le modèle Power BI.

# Chapitre 6

## Dashboards et Visualisations

### 6.1 Architecture des Tableaux de Bord

Nous avons conçu 5 dashboards spécialisés répondant chacun à des besoins analytiques distincts :

TABLE 6.1 – Structure des dashboards Power BI

Dashboard	Audience	KPIs principaux
Vue Exécutive	Direction générale	CA total, évolution YoY, marge globale
Analyse Clients & Fidélité	Marketing & CRM	RFM, lifetime value, performance par ville
Performance Produits	Direction commerciale	ABC, top ventes, matrice CA/Marge
Retours & Logistique	Supply Chain	Taux de retour, indice qualité, impact financier
Trafic Web & Conversion	E-commerce manager	Conversion, abandon panier, tunnel digital

### 6.2 Dashboard 1 : Vue Exécutive

#### 6.2.1 Objectif

Fournir à la direction générale une vision synthétique et instantanée de la performance commerciale globale pour une prise de décision rapide et éclairée.

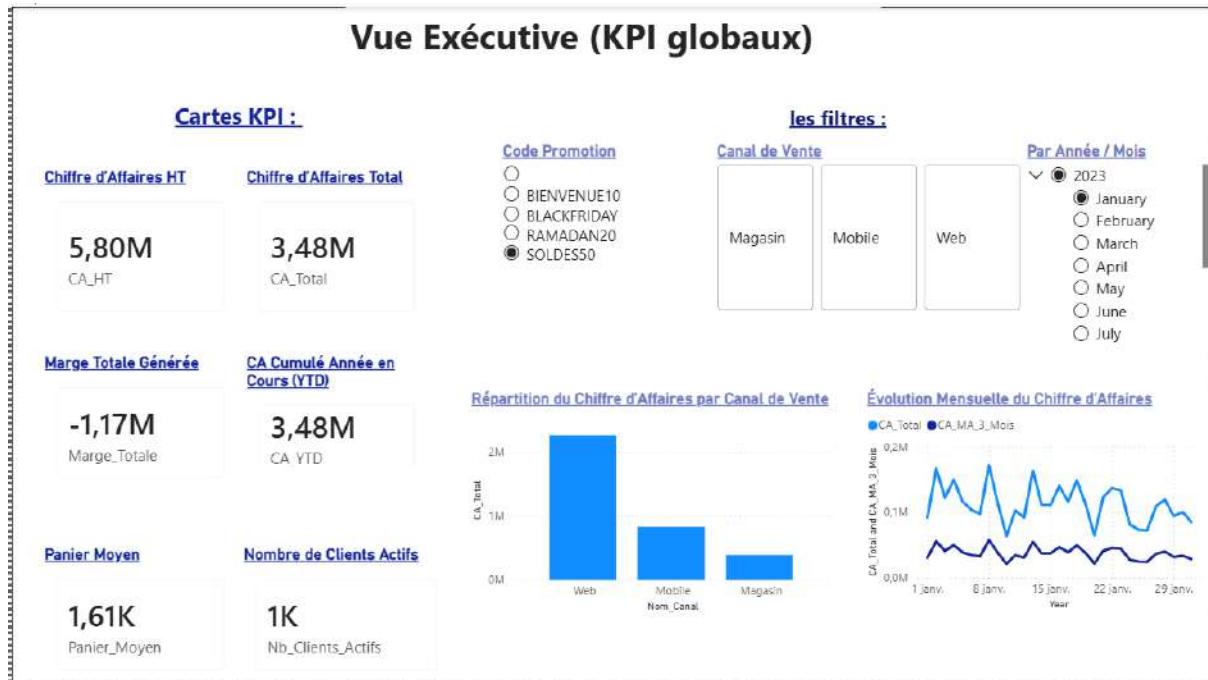


FIGURE 6.1 – Dashboard 1 - Vue Exécutive : KPI globaux et répartition par canal

### 6.2.2 Composantes Principales

- 6 cartes KPI : CA HT (5,80M MAD), CA Total (3,48M MAD), Marge (-1,17M MAD), CA YTD, Panier Moyen (1,61K MAD), Clients Actifs (1K)
- Répartition CA par canal : Web 57%, Mobile 29%, Magasin 14%
- Évolution mensuelle du CA avec moyenne mobile 3 mois
- Filtres : Code promotion, Canal, Période

### 6.2.3 Cas d'Usage

- **Réunion direction** : Vision hebdomadaire de la performance en 60 secondes
- **Analyse promotionnelle** : Comparaison rapide de l'efficacité des campagnes (SOLDES50, BLACKFRIDAY, RAMADAN20)
- **Pilotage omnicanal** : Allocation des investissements selon performance par canal

### 6.2.4 Valeur Ajoutée

Porte d'entrée du système décisionnel offrant une vue 360° (financière, temporelle, omnicanale) avec KPI immédiatement visibles et interface accessible sans expertise technique.

## 6.3 Dashboard 2 : Analyse Clients & Fidélité

### 6.3.1 Objectif

Permettre aux équipes Marketing et CRM d'analyser le comportement client, d'identifier les segments à forte valeur ajoutée et d'optimiser les stratégies de fidélisation et d'acquisition.



FIGURE 6.2 – Dashboard 2 - Analyse Clients & Fidélité : Segmentation RFM et performance par ville

### 6.3.2 Composantes Principales

- KPI : 1K clients actifs, CA par client 2,99K MAD, Panier moyen 1,66K MAD
- Segmentation RFM : Gold (625 clients, 81,64% du CA), Silver (83), Bronze (208), Nouveau (27)
- Treemap géographique : Casablanca et Rabat dominent, villes secondaires sous-exploitées
- Top 20 clients avec segment, CA et transactions
- Filtres : Ville, Segment RFM, Période

### 6.3.3 Cas d'Usage

- **Programme VIP** : Créer offres exclusives pour Gold (81% du CA)
- **Campagne fidélisation** : Faire évoluer clients Silver vers Gold
- **Réactivation** : Cibler Hassan Tazi (Casablanca) - CA élevé (29K MAD) mais segment Bronze
- **Expansion géographique** : Développer Agadir, El Jadida

### 6.3.4 Valeur Ajoutée

Segmentation actionnable basée sur RFM combinée à une vision géographique permettant de répondre à : qui sont nos meilleurs clients et où se trouvent-ils ? Identification rapide des segments nécessitant attention prioritaire.

## 6.4 Dashboard 3 : Performance Produits

### 6.4.1 Objectif

Permettre à la direction commerciale et aux category managers d'optimiser le catalogue produits, d'identifier les best-sellers et d'appliquer la classification ABC pour une gestion efficace des stocks et de l'assortiment.

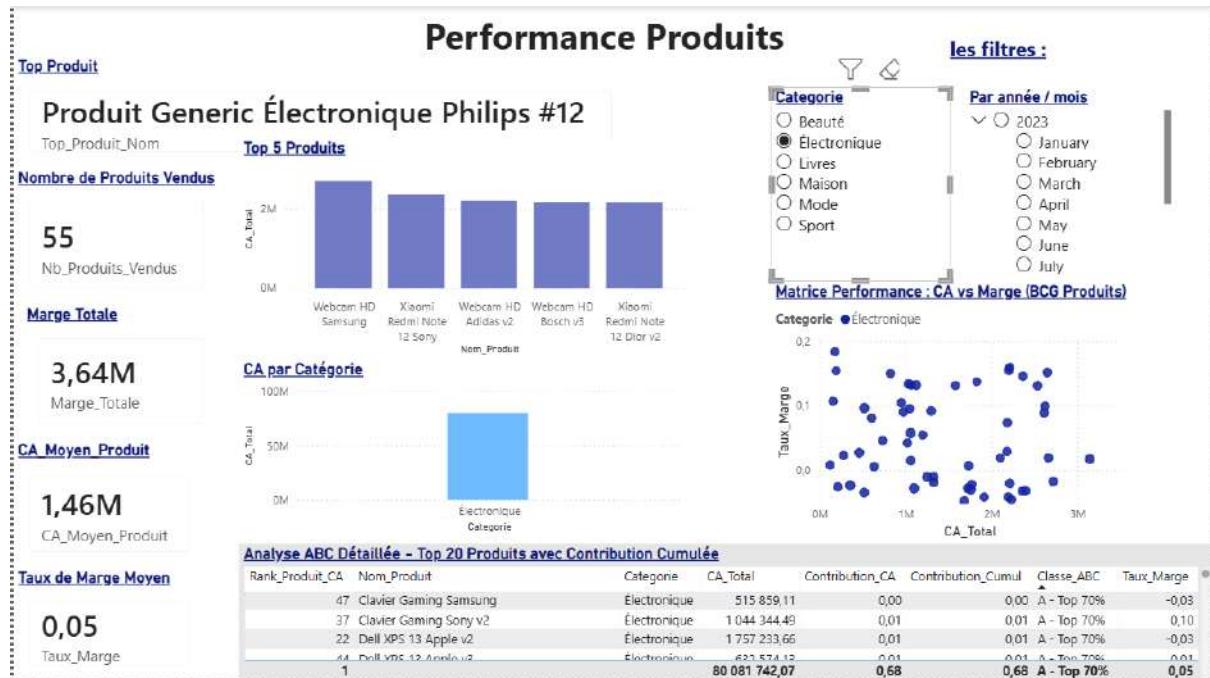


FIGURE 6.3 – Dashboard 3 - Performance Produits : Analyse ABC et contribution au CA

### 6.4.2 Composantes Principales

- KPI : Top produit (Philips #12), 300 références actives, Marge totale 5,71M MAD, CA moyen 393,79K MAD, Taux marge moyen 5%
- Top 5 produits : Webcam HD Samsung, Xiaomi Redmi Note 12 Sony (~2M MAD chacun)
- CA par catégorie : Électronique leader (80M MAD, 40% du CA total)
- Matrice CA vs Marge : Identification Stars, Cash Cows, Question Marks, Dogs
- Analyse ABC : Tableau 20 meilleurs produits avec contribution cumulée
- Filtres : Catégorie, Période

### 6.4.3 Cas d'Usage

- Optimisation catalogue : Déréférencer produits Classe C (< 10% CA)
- Gestion stocks : Prioriser réapprovisionnements Classe A
- Stratégie pricing : Ajuster prix produits forte marge/faible volume
- Négociation fournisseurs : Obtenir meilleures conditions sur top produits

#### 6.4.4 Valeur Ajoutée

Application rigoureuse de la loi de Pareto (80/20) avec double grille de lecture ABC et matrice CA/Marge pour optimiser simultanément chiffre d'affaires et rentabilité du catalogue.

### 6.5 Dashboard 4 : Retours & Logistique

#### 6.5.1 Objectif

Permettre aux équipes Supply Chain et Service Client d'analyser les retours produits, d'identifier les motifs de réclamation et de mesurer l'impact financier pour améliorer la qualité et réduire les coûts.

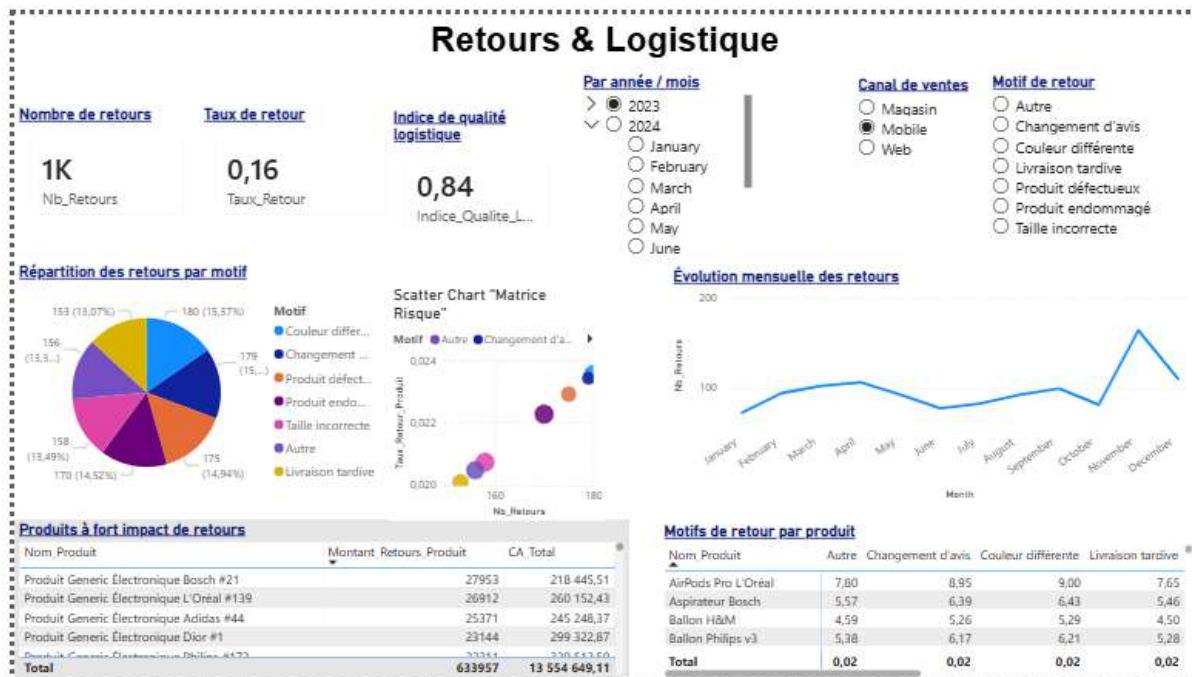


FIGURE 6.4 – Dashboard 4 - Retours & Logistique : Analyse des retours et indice qualité

#### 6.5.2 Composantes Principales

- KPI : 1K retours, Taux 16%, Indice qualité logistique 0,84
- Répartition par motif : Couleur différente 18,3%, Changement d'avis 18%, Taille incorrecte 17,2%, Livraison tardive 17%
- Scatter plot motif vs montant retours
- Évolution mensuelle : Pic octobre-novembre (200 retours)
- Tableau produits à fort impact : Total 634K MAD de retours sur 13,5M MAD de CA
- Analyse croisée produit/motif
- Filtres : Année/Mois, Canal, Motif de retour

### 6.5.3 Cas d'Usage

- **Amélioration fiches produits** : Réduire retours "couleur différente" et "taille incorrecte" (53% du total) via meilleurs visuels et guides des tailles
- **Contrôle qualité** : Audit produits Electronique (catégorie dominante dans retours)
- **Formation fournisseurs** : Partager données retours pour amélioration collaborative
- **Négociation logistique** : Réduire "livraison tardive" (17%)

### 6.5.4 Valeur Ajoutée

Vision complète de l'impact financier (634K MAD) avec identification des causes racines par motif. Réduction du taux de 16% à 10% représenterait une économie de 240K MAD annuels. Pilotage proactif satisfaction client et rentabilité.

## 6.6 Dashboard 5 : Trafic Web & Conversion

### 6.6.1 Objectif

Permettre au responsable e-commerce et aux équipes digitales d'analyser le comportement des visiteurs, d'évaluer l'efficacité du tunnel de conversion et d'identifier les points de friction dans le parcours d'achat pour maximiser les ventes en ligne.



FIGURE 6.5 – Dashboard 5 - Trafic Web & Conversion : Analyse du parcours client digital

### 6.6.2 Composantes Principales

- KPI : Taux conversion 50%, 50K sessions avec achat, 100K sessions totales, Taux abandon panier 15%, CA potentiel perdu 35,59M MAD

- Tunnel conversion : 100K sessions → 15K paniers (15,1%) → 50K achats
- Abandon panier par canal : Équilibré 33,33% (Magasin, Mobile, Web)
- Évolution mensuelle : Stable 10-15K abandons/mois
- Sessions avec achat par mois : Répartition équitable
- Filtres : Canal, Année, Mois

### 6.6.3 Cas d'Usage

- **Réduction abandon panier** : Implémenter relance email (récupération 10-15%), afficher frais livraison dès page produit, simplifier checkout
- **Optimisation conversion** : Audit UX/UI tunnel de conversion, ajouter options paiement (Cash on Delivery, fractionné)
- **A/B Testing** : Mesurer impact modifications sur taux conversion
- **Allocation budgets** : Calculer ROI campagnes retargeting sur paniers abandonnés

### 6.6.4 Valeur Ajoutée

Quantification précise du manque à gagner (35,59M MAD). Réduction du taux d'abandon de 15% à 10% générera un gain de 11,86M MAD annuels. Distribution équilibrée des abandons (33,33% par canal) indique problème transversal nécessitant actions structurelles (politique livraison, options paiement, processus checkout).

## 6.7 Interactivité Globale

### 6.7.1 Fonctionnalités Communes

- **Slicers** : Période, Canal, Région, Segment RFM, Catégorie (hiérarchique)
- **Drill-Through** : Top Produits → Détail produit | Segments RFM → Détail client
- **Tooltips** : Survol produit (évolution CA 30j, taux retour) | Survol ville (top 3 catégories, segment dominant)
- **Cross-Filtering** : Tous visuels interconnectés pour exploration dynamique
- **Synchronisation** : Filtres persistent entre dashboards

# Conclusion Générale

Ce projet a démontré la mise en œuvre complète d'une solution de Business Intelligence pour l'analyse e-commerce multi-canaux, couvrant l'intégralité de la chaîne décisionnelle.

À travers la conception d'un modèle dimensionnel en étoile, l'intégration de sources hétérogènes (SQL, Excel, CSV, JSON, XML), et la création de dashboards interactifs dans Power BI, nous avons développé un système analytique robuste répondant aux besoins métier de différentes parties prenantes.

Les principaux acquis de ce projet incluent :

- La maîtrise du processus ETL complet avec Power Query, incluant nettoyage, transformation et gestion de la qualité des données
- La modélisation analytique selon les bonnes pratiques BI (schéma en étoile, gestion des relations, optimisation des performances)
- Le développement de mesures DAX avancées (Time Intelligence, analyses RFM, classification ABC, indicateurs de conversion)
- La création de 5 dashboards spécialisés offrant des insights actionnables quantifiés (240K MAD d'économies potentielles sur retours, 11,86M MAD de gain possible via réduction abandon panier)

Les analyses réalisées ont permis d'identifier des leviers d'amélioration concrets : optimisation du catalogue produits via l'analyse ABC, ciblage des clients Gold représentant 81% du CA, réduction des retours concentrés sur trois motifs principaux (53% du total), et amélioration du tunnel de conversion digital.

Ce projet illustre comment une architecture BI bien conçue transforme des données brutes en intelligence stratégique, permettant une prise de décision éclairée et data-driven dans un contexte e-commerce compétitif.

## Perspectives d'évolution :

- Intégration de modèles prédictifs (churn client, forecasting ventes)
- Automatisation du rafraîchissement des données via Power BI Service
- Enrichissement avec données externes (réseaux sociaux, météo, concurrence)
- Déploiement d'alertes automatiques sur KPIs critiques