

Annexe E – Dictionnaire des données

NovaRetail

Ce document décrit où trouver les données du cours, comment les installer, et le contenu exact de chaque fichier. Les fichiers se trouvent dans le dossier **donnees/** fourni avec la formation.

1. Où sont les données et comment les installer ?

Les jeux de données sont livrés avec le cours, dans le sous-dossier **donnees/** :

```
cours-powerbi/
├── donnees/                ← LES DONNEES DES TP (indispensables)
│   ├── ventes_2023.csv
│   ├── ventes_2024.csv
│   ├── produits.csv
│   ├── magasins.csv
│   ├── clients.csv
│   └── budget_2024.csv
├── module-01 ... module-07
└── annexes/
```

Installation avant le module 2

1. Créez votre dossier de travail : **C:\Formations\PowerBI\NovaRetail**
2. Copiez-y **tout le dossier donnees/** (les 6 fichiers)
3. Vous devez obtenir :

```
C:\Formations\PowerBI\NovaRetail\
├── donnees\
│   ├── ventes_2023.csv
│   ├── ventes_2024.csv
│   ├── produits.csv
│   ├── magasins.csv
│   ├── clients.csv
│   └── budget_2024.csv
```

Important : ces fichiers ne s'ouvrent pas « pour rien ». Ils sont destinés à être connectés dans Power BI Desktop via **Accueil > Obtenir les données > Texte/CSV**, exactement comme le montrent les modules 2, 3 et 4. Sans eux, aucun exercice pratique n'est réalisable.

Caractéristiques techniques communes : encodage **UTF-8**, séparateur **point-virgule (;)**, dates au format **ISO AAAA-MM-JJ**, nombres avec point décimal.

2. Les 6 fichiers en détail

2.1 **produits.csv** — Dimension Produit (21 lignes)

Colonne	Type	Description
IDProduit	Entier	Clé primaire (1000 à 1020)
NomProduit	Texte	Libellé commercial
Categorie	Texte	Informatique, Téléphonie, Audio & Image, Électroménager
CoutUnitaire	Entier	Coût d'achat en € (base du calcul de marge)
PrixCatalogue	Entier	Prix public avant remise

2.2 **magasins.csv** — Dimension Magasins (13 lignes)

Colonne	Type	Description
IDMagasin	Entier	Clé primaire ; 20 = le canal Web (Novaretail.com)
NomMagasin	Texte	Nom de l'enseigne
Ville	Texte	Ville du magasin (- pour le Web)
Region	Texte	8 régions françaises (+ National pour le Web)
Canal	Texte	Physique ou Web
SurfaceM2	Entier	Surface de vente (0 pour le Web)

2.3 **clients.csv** — Dimension Clients (80 lignes)

Colonne	Type	Description
IDClient	Entier	Clé primaire (1 à 80)
NomClient	Texte	Nom fictif
Segment	Texte	Particulier, Professionnel, VIP
Ville	Texte	Ville de résidence
DateInscription	Date	Date d'entrée en base (2019-2025)

2.4 **budget_2024.csv** — Budget par catégorie (4 lignes × 12 mois)

Format **large** exprès : les mois sont en colonnes. Il servira au module 4 pour apprendre à **décroisser** un tableau (Unpivot) et obtenir 48 lignes propres.

Colonne	Type	Description
Categorie	Texte	Les 4 catégories de produits

Colonne	Type	Description
Janvier ... Décembre	Entier	Budget de CA mensuel visé en €

2.5 ventes_2023.csv — Table de faits, année 2023 (907 lignes)

2.6 ventes_2024.csv — Table de faits, année 2024 (1050 lignes)

Colonne	Type	Description
IDVente	Entier	Identifiant ligne de vente (100000+ pour 2023, 200000+ pour 2024)
Date	Date	Date de la vente
IDProduit	Entier	Clé étrangère vers produits
IDMagasin	Entier	Clé étrangère vers magasins
IDClient	Entier	Clé étrangère vers clients (peut être vide !)
Quantite	Entier	Quantité vendue
PrixAchat	Décimal	Prix unitaire réellement facturé (après remise)

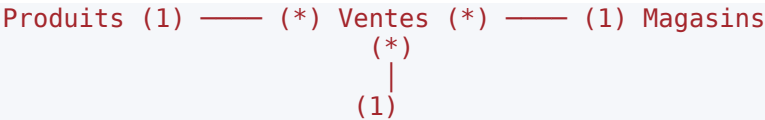
3. Anomalies VOLONTAIRES dans ventes_2023.csv

Le fichier 2023 est **pollué exprès** pour les exercices de nettoyage du module 3. Ne soyez pas surpris :

Anomalie	Nombre de lignes concernées	Exercice associé (module 3)
Espaces parasites autour du prix (" 89 ")	10	Trimmer (Format > Trim)
Quantité saisie en toutes lettres (deux)	6	Remplacer les valeurs puis retyper
IDClient vide (vente anonyme)	5	Statistiques + décision métier
Lignes entièrement dupliquées	7	Supprimer les doublons

Après nettoyage complet, vous devez obtenir **900 lignes propres**. Le fichier 2024, lui, est irréprochable : il sert de référence.

4. Schéma relationnel attendu (module 5)



DimDate (à créer : module 5)
Budget décroissé (1) — (*) via Categorie + Mois (relation composite, optionnel)

Volumes finaux attendus après tous les modules : **FactVentes = 1950 lignes** • DimDate = 731 lignes (2023-2024) • Budget décroissé = 48 lignes.

5. FAQ données

« **Je n'ai pas reçu le dossier donnees/** » : demandez-le au formateur ; il est aussi dans l'archive ZIP du cours ([cours-powerbi/donnees/](#)).

« **Puis-je utiliser mes propres fichiers ?** » : oui dès le module 6 (exercice personnalisé), mais faites d'abord tous les TP avec NovaRetail : les corrigés s'y réfèrent précisément.

« **Les chiffres sont-ils réalistes ?** » : les volumes et saisonnalité sont calibrés pour ressembler à un vrai SI de retail (pic novembre-décembre), tout en restant légers pour tourner sur n'importe quel poste.