

Projet data visualisation

Vatin Antoine

Business Intelligence Analyst

Mars 2025

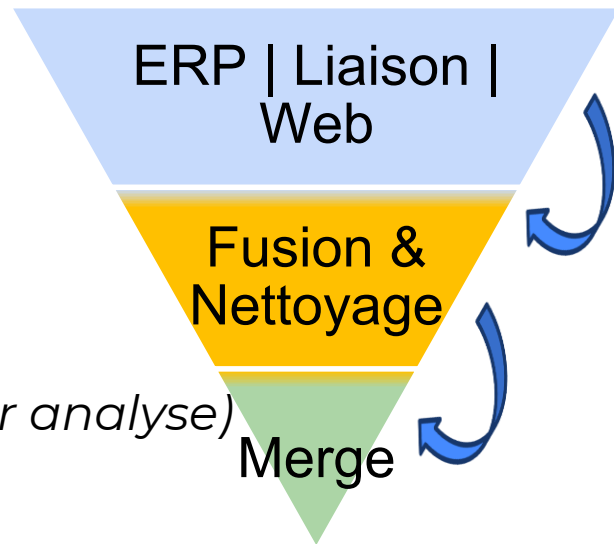
Analyses Exploratoires des Données

Elements disponibles :

- 3 Datasets : (ERP, Web et Liaison)
- Caractéristiques :
 - ERP : 825 observations, 6 colonnes
 - Web : 1513 observations, 29 colonnes
 - Liaison : 825 observations, 2 colonnes

Finalité - Harmonisation des données

- 1 Dataset : Merge (dataset consolidé et prêt pour analyse)
- Caractéristiques :
 - Merge : 714 observations, 18 colonnes
 - Conformité à la réglementation RGPD

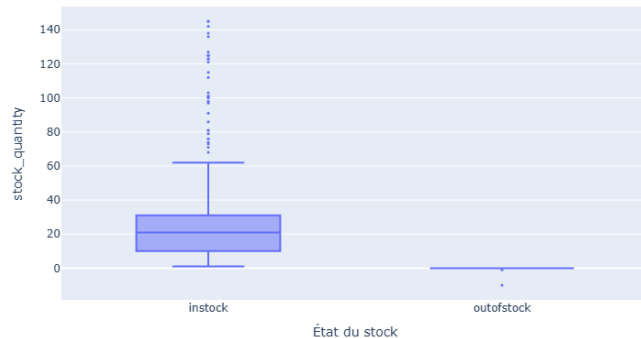


Analyses Exploratoires des Données

Identification des incohérences et préparation des données pour une analyse fiable

- *Inspection et validation des variables*
 - Vérification de la répartition des prix et des stocks
 - Détection des liens de corrélation entre les colonnes
- *Correction des valeurs aberrantes et incohérences*
 - **Stock et prix négatifs**
 - **Valeurs manquantes** (product_type, SKU)

Distribution de stock_quantity selon stock_status



Boite à moustache montrant la relation entre la quantité de stock et son statut

product_id	onsale_web	price
4233	0	-20.0
5017	0	-8.0
6594	0	-9.1

Exemple d'incohérences de prix

product_id	stock_quantity	stock_status
4973	-10	outofstock
5700	-1	outofstock
4039	3	outofstock

Exemple d'incohérences de stock

Analyses Exploratoires des Données

Traitement réalisés

- *Traitement des doublons et uniformisation*
 - **Suppression des SKU en doublons**
 - **Consolidation des informations** pour éviter les biais
- *Normalisation des données*
 - **Correction des statuts incohérents** entre quantité de stock et statut affiché

sku	product_type	post_type
13127-1	Vin	product
bon-cadeau-25-euros	Autre	attachment
13127-1	Vin	attachment
bon-cadeau-25-euros	NaN	product

Exemple de type de produit manquant et de doublon

product_id	price	stock_quantity	stock_status
3847	24.2	16	instock
3849	34.3	10	instock
3850	20.8	0	outofstock
4032	14.1	26	instock
4039	46.0	3	instock

Extraction à la suite de la correction des incohérences de statut de stock

Analyses Exploratoires des Données

🔧 *Création de nouvelles variables pour une meilleure analyse :*



Taux de marge



Mois de stock



Rotation des stocks



Prix HT

sku	taux_marge
16148	69.753086
15415	64.670659
14864	59.597258

sku	mois_stock	rotation_stock
11641	375.000000	0.032000
14923	331.200000	0.036232
12585	324.000000	0.037037

sku	prix_ht
16148	8.250000
15415	8.250000
14864	12.416667

Extrait de données enrichies

Analyses Exploratoires des Données

Problèmes détectés :

Erreurs de saisie :

- **Prix de vente négatifs (ERP)**
- **Incohérence de statut (ERP)**
- **Stock négatif (ERP)**
- **SKU manquante (Web)**

De calcul :

- **Marge négative**

De jointure :

- **'product_id' sans correspondance 'web_id' (liaison)**

De doublons :

- **Colonne 'post_type' ayant 'attachment' est un doublon inutile de 'product'**

De type :

- **Type de produit NaN pour « bon-cadeau-25-euros »**

Fusion ou consolidations des données

Choix des attributs

- *Seules les colonnes et lignes nous permettant d'effectuer nos analyses seront gardés*

Clés utilisées

- *id_web (Table liaison)*
- *product_id (Table erp)*
- *Sku (Table web)*

Nom de la colonne	Type de donnée	Nombre d'éléments non-nuls
product_id	int64	714
price	float64	714
stock_quantity	int64	714
stock_status	object	714
purchase_price	float64	714
sku	object	714
total_sales	float64	714
product_type	object	714
post_title	object	714

Liste des colonnes présentes dans la table prête pour les analyses et suppression des « lignes vides » (85)

Fusion ou consolidations des données

Vigilances particulières

- *Identification des lignes ne matchant pas*

product_id	price	stock_quantity	stock_status	purchase_price	id_web
4055	86.1	0	outofstock	37.88	NaN
4090	73.0	0	outofstock	33.79	NaN
4092	47.0	0	outofstock	25.25	NaN
4195	14.1	0	outofstock	7.36	NaN
4209	73.5	0	outofstock	33.01	NaN

Nombre de lignes sans correspondance entre erp et liaison : 91

Difficultés ou pièges rencontrés

- *Détermination que 'sku' correspondait à 'id_web' après analyse des structures de données.*

product_id	id_web	sku
5794	14692	14692
5827	15328	15328
4964	16515	16515
4223	16585	16585
5900	12869	12869

Extraction montrant le match entre 'id_web' et 'sku'

Analyses univariées du prix

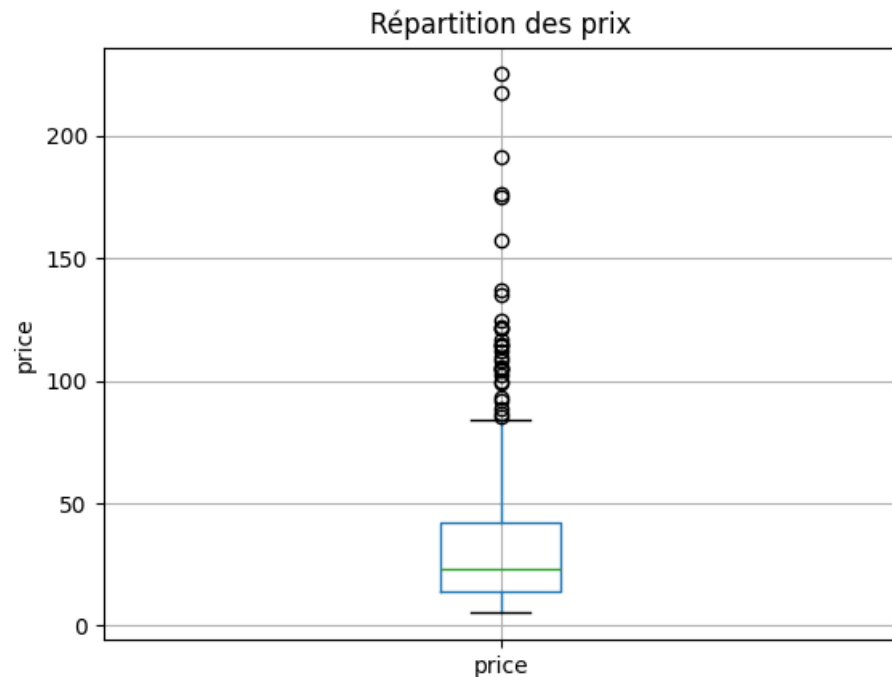


Analyse de la répartition des prix

- Ce *Boxplot* (boite à moustaches) met en évidence une distribution **asymétrique** avec une **forte présence de valeurs aberrantes (outliers)**, ce qui peut impacter les analyses

Limites de l'analyse

- Présence de nombreuses valeurs aberrantes
- Absence d'informations contextuelles
- Analyse statique sur une période donnée



Boite à moustaches représentant la répartition des prix

Analyses univariées du prix

La méthode du Z-score

- **Permet d'identifier les valeurs extrêmes** (outliers) en calculant combien d'écarts-types une valeur s'éloigne de la moyenne.
- Forte présences d'**outliers** (13), ces valeurs peuvent correspondre à des produits de luxe ou des erreurs de saisie.

Limites éventuelles de l'analyse

- Influence des valeurs extrêmes sur la moyenne (seuil prix de détermination des outliers est de 115,12€)

La formule utilisée est :

$$Z = \frac{X - \mu}{\sigma}$$

où :

- X est la valeur observée (le prix),
- μ est la moyenne du prix (ici 32.33),
- σ est l'écart-type du prix (ici 27.60).

Z-score minimum : -0.98

Z-score maximum : 6.98

Nombre de valeurs aberrantes ($|z\text{-score}| > 3$) : 13

Résultat du calcul du Z-score, montrant le nombre de valeurs extrêmes

Analyses univariées du prix

Identification par l'interval interquartile (IQR)

- La méthode **IQR détecte plus d'outliers** (31) que la méthode du Z-score (13), car la distribution **des prix est asymétrique**

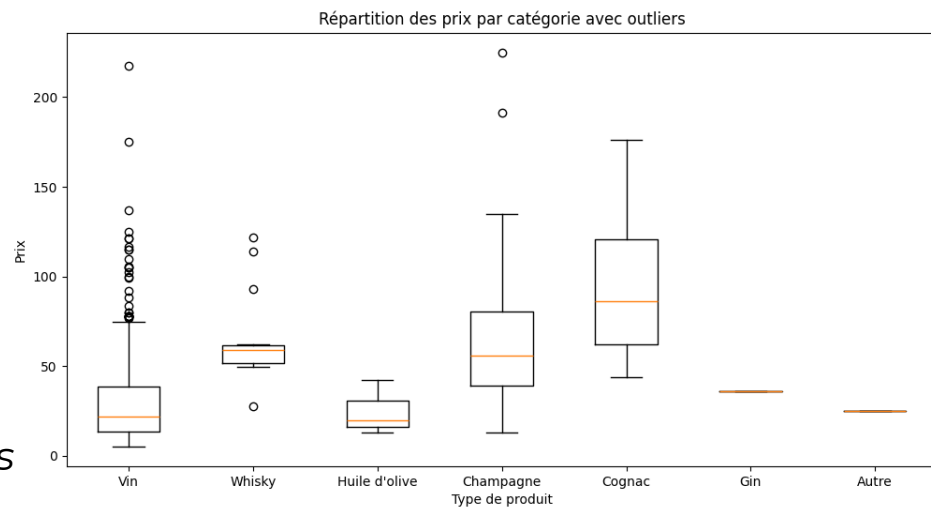
Limites éventuelles de l'analyse

- La méthode est **sensible aux variations de données**, si le nombre d'observations est faible l'IQR peut ne pas être fiable
- Exclusion potentielle de données pertinentes** en fixant un seuil rigide, en particulier dans des marchés où les écarts de prix sont normaux.

$$IQR = Q_3 - Q_1$$

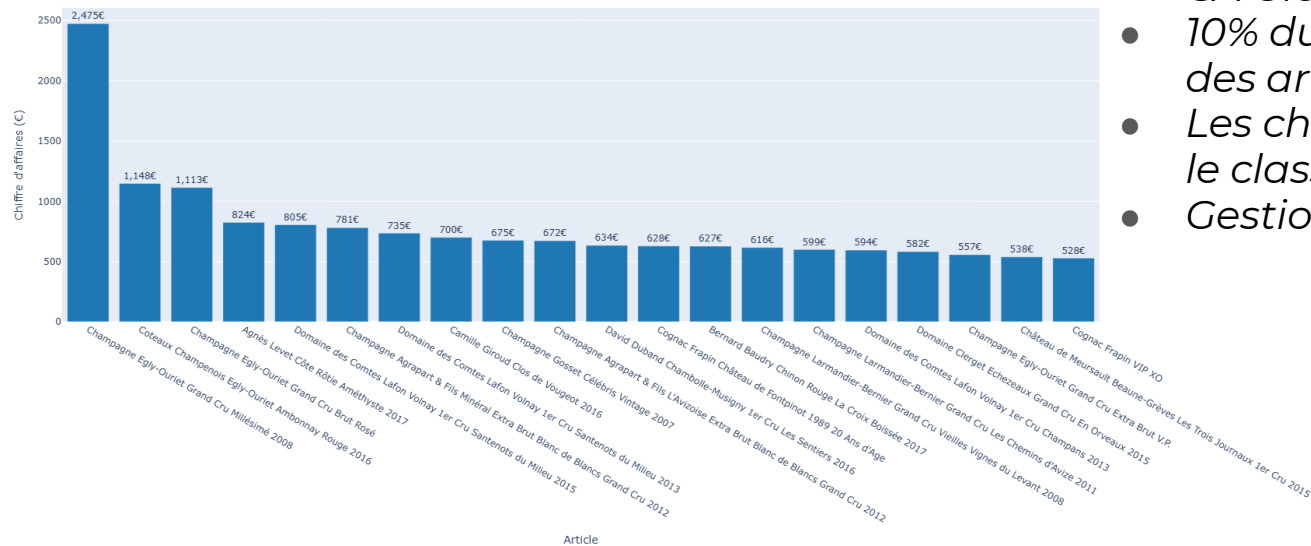
L'IQR (Intervalle Interquartile) est défini comme la différence entre le troisième quartile (Q_3) et le premier quartile (Q_1) d'une distribution. Il représente l'intervalle dans lequel se trouve la moitié centrale des données.

Une valeur x est considérée comme aberrante si $x < Q_1 - 1.5 \times IQR$ ou $x > Q_3 + 1.5 \times IQR$



Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

Top 20 des articles en chiffre d'affaires



Top 20 des articles en CA

- CA Global de 143 680€
- 10% du CA représenté par le Top 20 des articles
- Les champagnes et vins dominent le classement
- Gestion des stocks primordial

Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

—

Calcul des 20/80

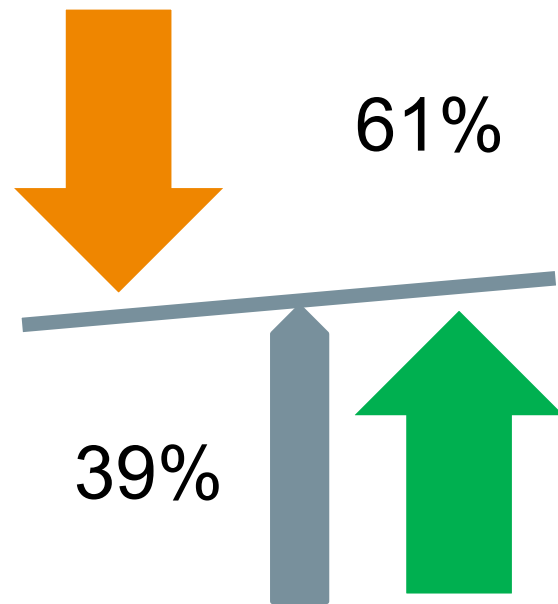
- Définition : **20%** des produits génèrent **80%** du chiffre d'affaires total

Chez Bottleneck

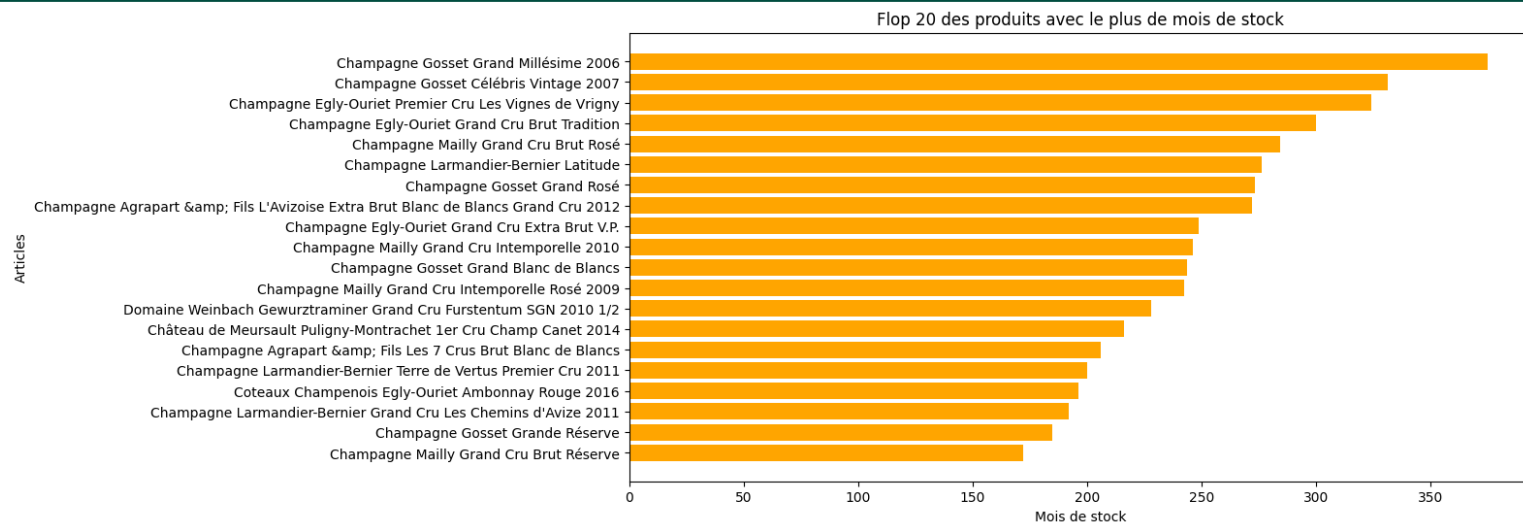
- **434** articles représentent **80%** du CA
- Soit environ **61%** des produits

Optimisation du portefeuille produits

- **39%** des références absorbent des ressources sans générer une valeur proportionnelle.
- Une concentration sur les **434 articles stratégiques** permettrait d'optimiser la rentabilité et la gestion des stocks.



Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



Immobilisation de stock critique : jusqu'à 25 ans !



Optimisation du portefeuille produits, impact financier : **87 108,25€**



Envisager une **suppression** progressive



Stratégie marketing et commerciale



Mettre en avant ces produits

Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



Le stock total : 16 739 unités dont une part non rentable

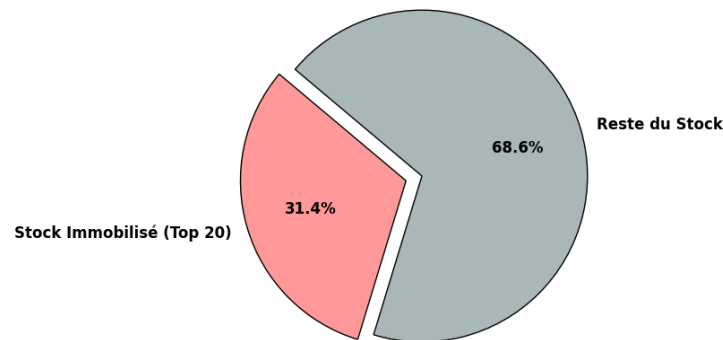
Immobilisation financière totale : 277 305€

- *Rotation insuffisante sur certains produits*
- *Coût financier et logistique*
- *Surstockage vs. Rupture*

Recommandations

- ✂ *Segmentation du stock*
- 📊 *Analyse de la demande*
- ⚠ *Plan de réduction des stocks dormants*
- 🔄 *Optimisation des réapprovisionnements*

Part du Stock Immobilisé dans la Valorisation Totale



Stock Immobilisé : 87 108,25 €

Taux de Marge et Rentabilité des Produits

- Taux de marge minimum : **-86,39 %**
- Taux de marge maximum : **91,41 %**
- Taux de marge minimum sur les produits rentables : **30,73 %**
- Un produit (Champagne Egly-Ouriet Grand Cru Blanc de Noirs) affiche un taux de marge négatif **perte potentielle 6 493€**

Recommandations

- 📌 Optimisation des prix de vente
- 📊 Segmentation des produits par rentabilité
- 🚀 Plan d'action pour la référence déficitaire
- ⚖️ Alignement entre les coûts et la politique tarifaire
- 💰 L'optimisation des marges est un levier clé pour améliorer la rentabilité globale de Bottleneck et sécuriser notre trésorerie



Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

—

Analyse des Corrélations


- *Corrélation positive entre le stock et les ventes*
- *Corrélation négative entre le prix et les ventes*
- *Corrélation légèrement négative entre le prix et le stock*

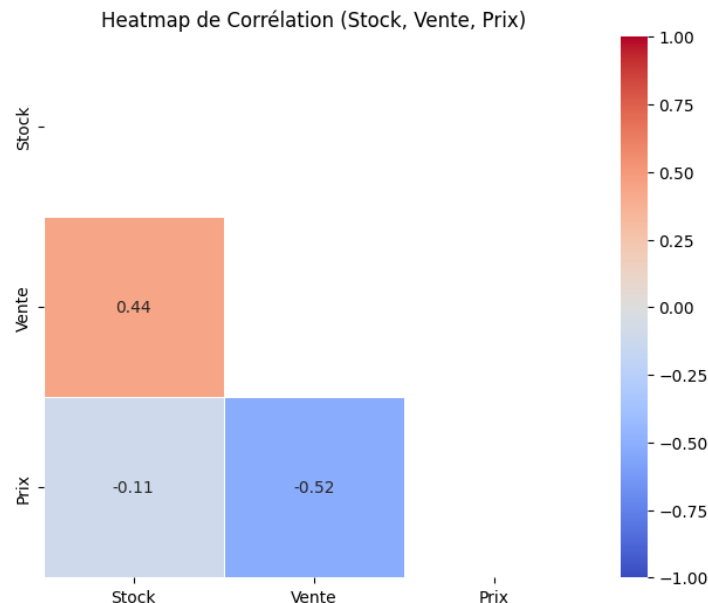
Recommandations

 *Optimisation des niveaux de stock*

 *Segmentation des stocks et gestion différenciée*





 *Stratégie tarifaire dynamique*

 *Analyse plus fine des produits premium*






Actions pour la suite




Court terme (0-3 mois) :

-  **Plan d'écoulement des stocks dormants** → Accélération de la rotation des stocks immobilisés
-  **Ajustement des prix** des références déficitaires
-  **Lancement d'actions marketing** sur les produits stratégiques
-  **Tableau de bord & KPI clés** (taux de rotation, marge, stock immobilisé)

Moyen terme (3-6 mois) :

-  **Stratégie de pricing dynamique** & ajustement des réapprovisionnements
-  **Segmentation des produits** selon la rentabilité
-  **Suivi des ventes & ajustement des stratégies marketing**

Long terme (6-12 mois) :

-  **Automatisation des processus** pour la gestion des stocks
-  **Définition des seuils critiques** (ex. +12 mois sans vente = action immédiate)
-  **Optimisation des relations fournisseurs**

Point sur les compétences apprises

Points positifs : Ce qui s'est bien passé


- ✓ **Analyses fluides** et **structurées**
- ✓ **Détection intuitive des erreurs** dans un DataFrame (`df.info()`, `df.isnull().sum()`, `describe()`)
- ✓ **Recherche de solutions** stimulante malgré sa complexité

Défis rencontrés : Ce qui a été le plus difficile

- L'intégration des **notions de jointures** et leur impact sur la **perte de données** (qu'elle soit voulue ou subie) a été un point particulièrement complexe à appréhender. (Fusion de **df_merge** avec **df_web**)
- Comprendre comment minimiser la perte d'information lors des **fusions de tables** reste un **défi technique** à surmonter.

Point sur les compétences apprises

Axes d'amélioration : Points à approfondir

 **Compréhension des jointures** (tester différents types de merge et observer leur impact sur les données)

 **Gestion des pertes de données** utiliser **.shape** avant et après une jointure pour quantifier les pertes

 **Anticipation des impacts**

 **Formation continue** (ex : documentation Pandas, cours SQL, etc.)

Ces axes d'amélioration guideront mes prochaines analyses pour fiabiliser la consolidation des données et optimiser leur exploitation dans la prise de décision.