[Bottleneck]



Huang Nicolas
Déveloper IA
02/10/2025

Analyses Exploratoires des Données

• Datasets utilisés :

- **ERP**: informations produits internes (prix d'achat, stock, taux de marge...)
- Web: ventes e-commerce (SKU, prix, quantités vendues)
- **Liaison**: correspondance SKU ↔ ERP

• Traitements réalisés :

- Normalisation des clés de jointure (ref_web_norm, ref_erp_norm)
- Conversion et typage des colonnes numériques (prix, quantités, stock...)
- Jointure des sources Web ↔ ERP via la table de liaison

• Nettoyages & enrichissements:

- Détection des valeurs aberrantes (méthodes Z-Score et IQR)
- Suppression / correction des incohérences (prix négatifs, références vides)
- Création de nouvelles métriques : Chiffre d'affaires, marge, mois de stock, rotation

• Points d'attention :

- Données manquantes ou incohérentes (SKU non reliés, duplications)
- Unités de stock hétérogènes selon les références
- Importance de fiabiliser les clés de jointure avant intégration ERP

Fusion ou consolidations des données

• Choix des attributs :

- Références produits normalisées (ref_web_norm, ref_erp_norm)
- Prix unitaire (Web vs ERP)
- Quantités vendues (Web)
- Stock disponible (ERP)
- Taux de marge et prix d'achat (ERP)

• Clés utilisées :

- Jointure basée sur la table de liaison Web ↔ ERP
- Normalisation des identifiants produits (suppression espaces, zéros, doublons)
- Vigilances particulières :
- Références manquantes ou incohérentes entre ERP et Web
- Doublons dans les ventes Web (SKU identiques sous plusieurs formats)
- Alignement des unités (ex. conditionnements, arrondis de prix)

• Difficultés / pièges rencontrés :

- Gestion des références sans correspondance ERP
- Prix aberrants (valeurs négatives, très éloignées de la médiane)
- Stocks hétérogènes (valeurs nulles, anomalies sur seuils critiques)
- Risque de biais lors du calcul du CA si données non nettoyées

Analyses univariées du prix

• Méthodes statistiques employées :

- Distribution des prix (moyenne, médiane, écart-type)

Détection des valeurs aberrantes via :

- Z-Score (|z| > 3)
- Écart interquartile (IQR) (valeurs hors [Q1 1.5×IQR ; Q3 + 1.5×IQR])

• Graphique + commentaires :

- Boxplot des prix → met en évidence plusieurs valeurs aberrantes (prix négatifs ou très élevés).
- Histogramme des prix → distribution asymétrique avec concentration entre X et Y € (majorité des produits).
- Outliers confirmés comme erreurs de saisie et non comme produits premium légitimes.

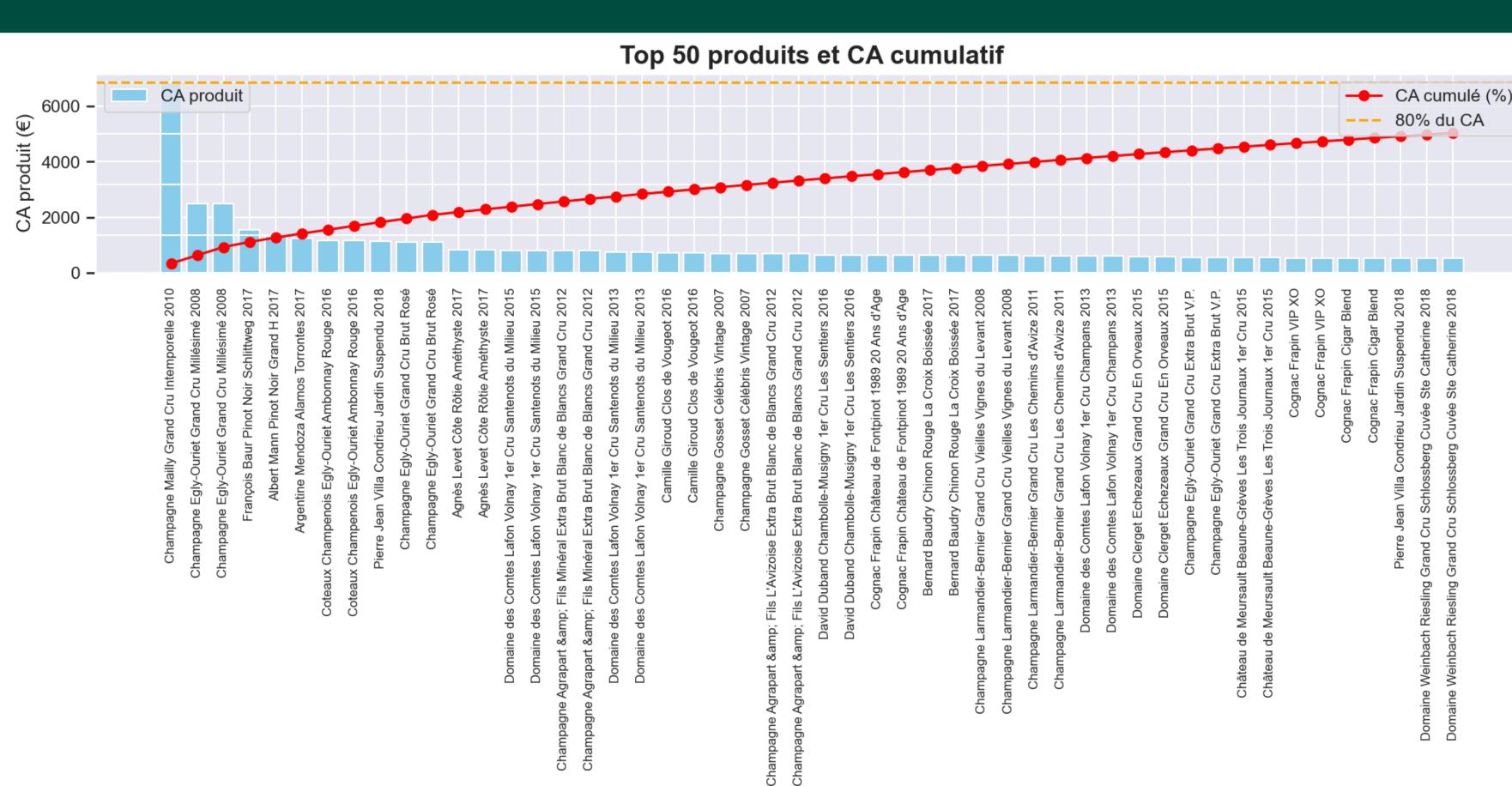
• Limites de l'analyse :

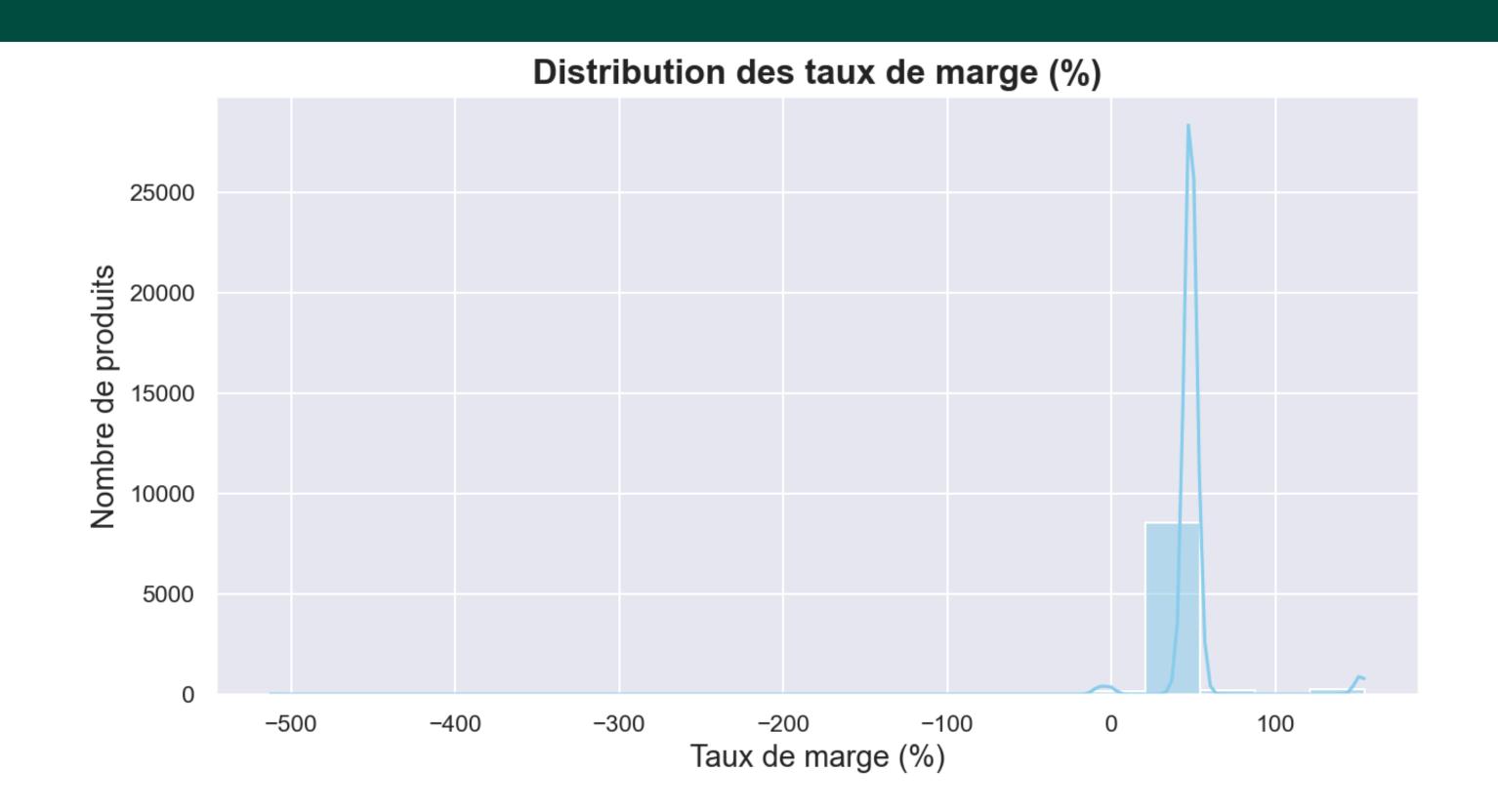
- Sensible aux doublons ou références Web sans correspondance ERP.
- Certaines anomalies nécessitent une validation métier (ex. prix promotionnels vs. erreurs de saisie).
- Analyse univariée limitée → ne tient pas compte de la relation prix ↔ ventes ou prix ↔ marge.

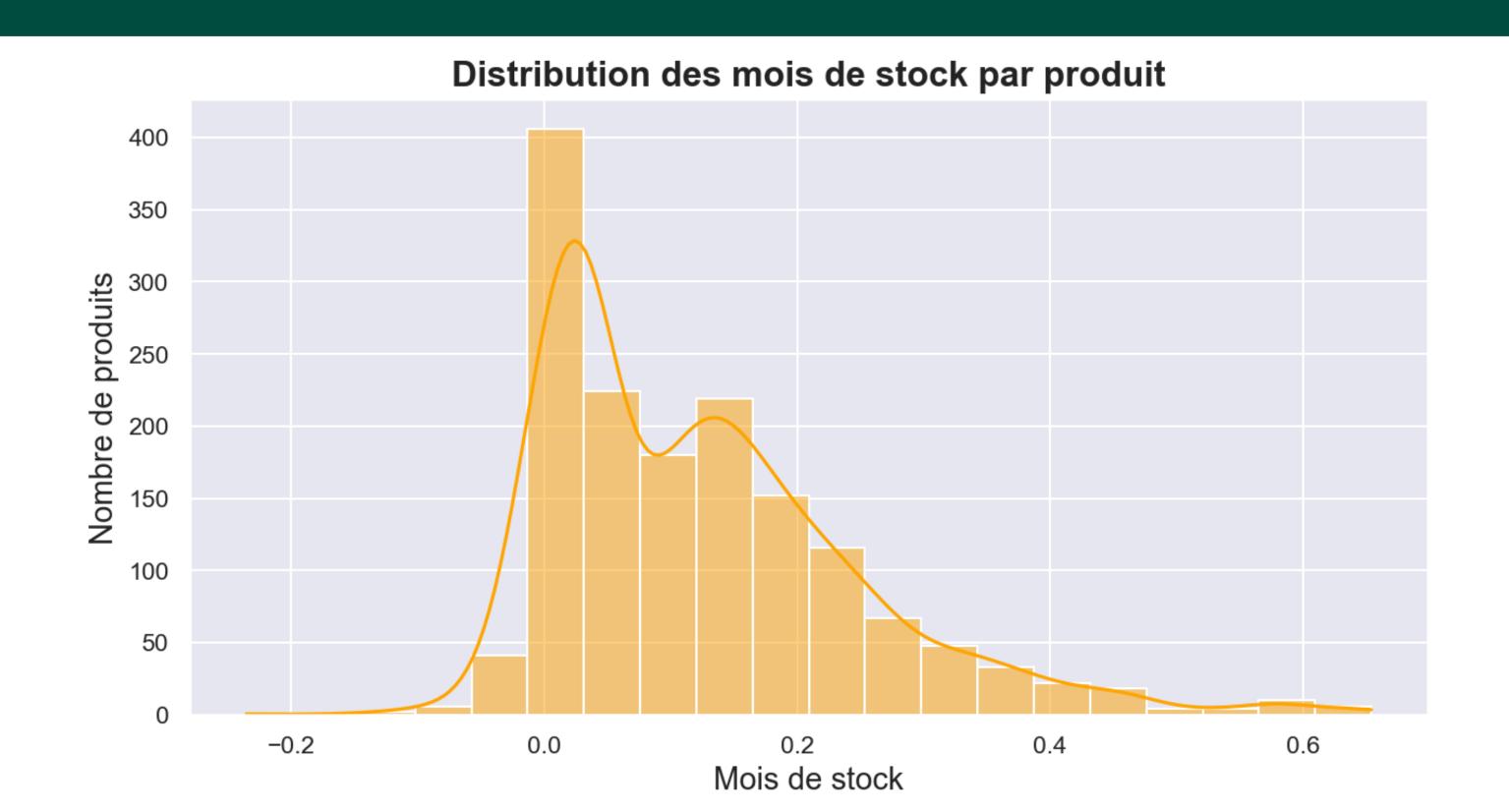
Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

- Méthodes statistiques employées :
 - Calcul du Chiffre d'Affaires (CA) par produit et global
 - Identification des produits stratégiques via la règle 20/80 (Pareto)
 - Analyse de la rotation des stocks (mois de stock, risque de rupture)
 - Calcul du taux de marge : (prix coût d'achat) / prix
 - Matrice de corrélations entre variables clés (prix, stock, CA, marge)

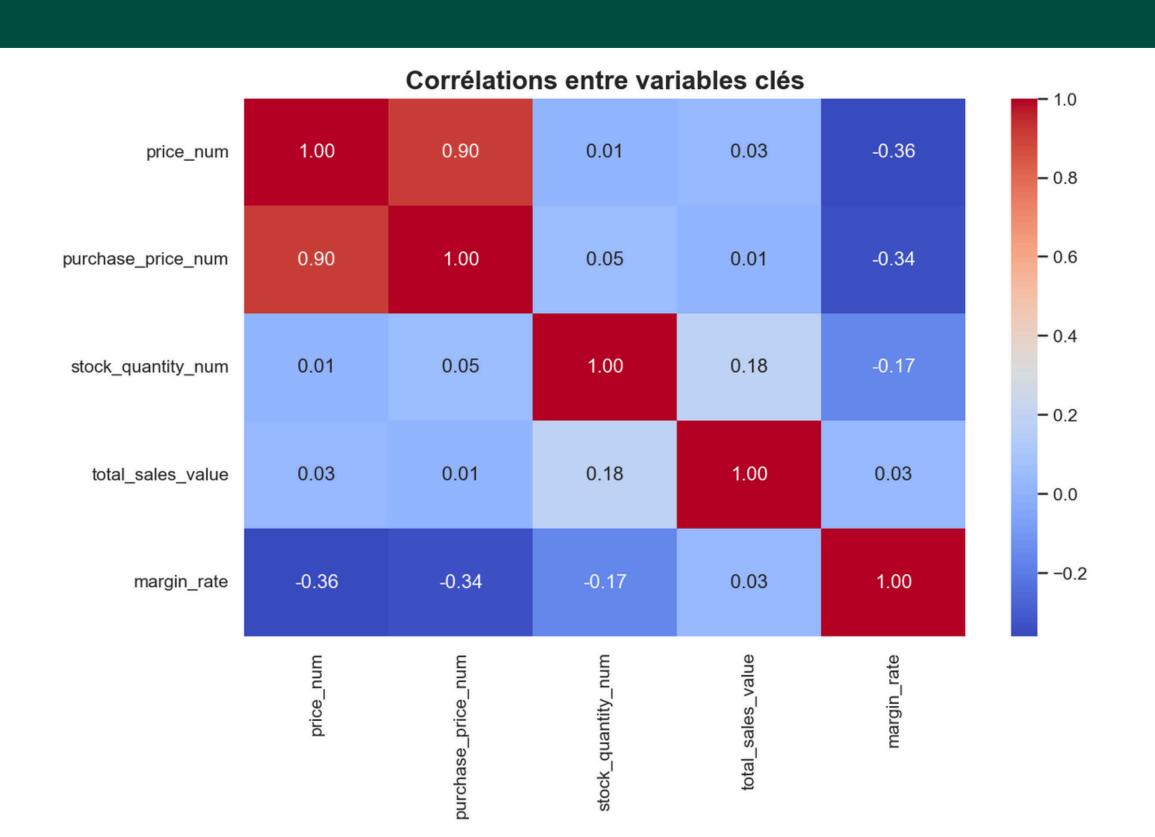
Analyses complémentaires CA, quantités, stocks, taux de marge et correlations







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



Actions pour la suite

• 1. Consolidation des données

- Finaliser la liaison Web ↔ ERP pour réduire les produits sans correspondance
- Mettre en place des contrôles automatiques sur les prix aberrants et les stocks incohérents

• 2. Amélioration des analyses

- Suivi régulier du 20/80 CA pour ajuster les priorités produits
- Mise en place d'un reporting automatisé (tableaux de bord Power BI / Python)
- Approfondir l'analyse des marges par catégorie de produits

• 3. Mise en œuvre opérationnelle ERP

- Intégration progressive des données propres dans l'ERP
- Formation des équipes aux nouveaux process de saisie et suivi
- Définir des KPI de pilotage (CA, marge, rotation stock, ruptures)

Point sur les compétences apprises

• Ce qui s'est bien passé

- Mise en place d'un pipeline clair de nettoyage et consolidation des données (ERP + Web)
- Application de méthodes statistiques (Z-score, IQR, corrélations) pour identifier les anomalies
- Génération automatique de graphiques et exports pour le support CODIR

• <u>A</u> Les principales difficultés rencontrées

- Gestion des valeurs manquantes et incohérentes dans les fichiers sources
- Normalisation des clés d'identifiants (références produit Web/ERP)
- Choix de représentations visuelles lisibles pour des volumes importants (ex. top produits CA)

- Les techniques avancées de feature engineering pour enrichir l'analyse
- L'interprétation fine des corrélations et causalités entre variables
- La conception de dataviz adaptées au CODIR (clarté, storytelling visuel)