

Optimisation de la Gestion Commerciale et des Stocks pour **bottleneck**

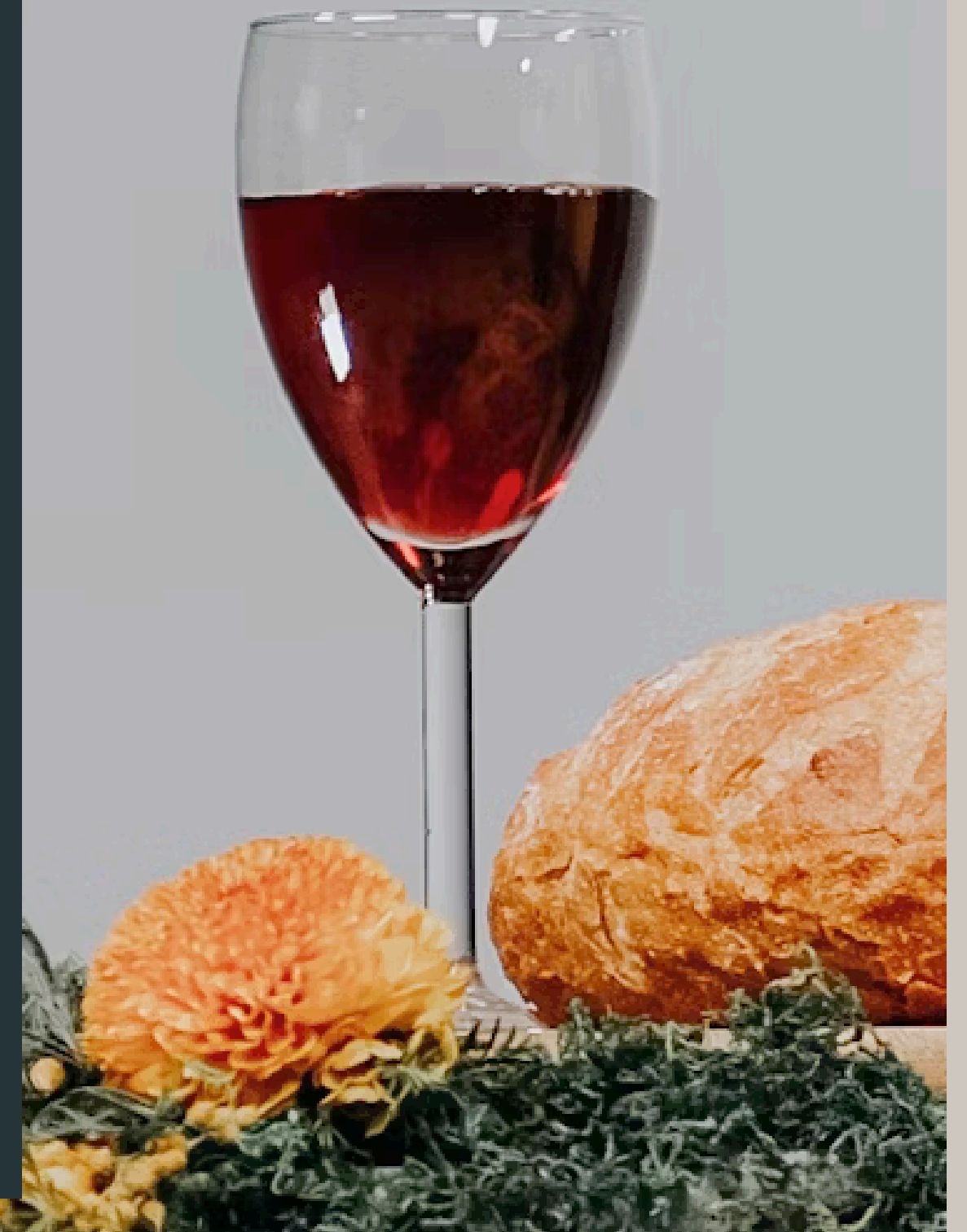
Présentation des résultats d'analyse des données

CONTEXTE ET OBJECTIFS:

Nous avons nettoyé les données de l'ERP et du site web pour obtenir un jeu des données cohérent et fiable.

L'objectif est de fournir des indicateurs clés à la direction et de poser les bases d'un projet de data visualisation.

Giulia Governatori





Nettoyage et Consolidation des Données

MÉTHODOLOGIE:

- Analyse exploratoire des fichiers pour identifier leur structure et détecter d'éventuelles anomalies
- Correction des incohérences
- Fusion des 3 fichiers (ERP, WEB, LIAISON)
- Nettoyage de la table agrégée.



SUPPRESSION DES DOUBLONS

📌 Problème identifié

Lors de l'analyse des données, 714 lignes avec des SKU dupliqués ont été détectées, alors qu'une clé primaire doit être unique dans une table.

📌 Solution appliquée

Un script Python a été mis en place pour fusionner ces doublons en appliquant des règles spécifiques:

- Si les valeurs sont identiques, elles sont conservées.
- Si ce sont des nombres divergents, ils sont additionnés (TOTAL_SALES).
- Si ce sont des textes divergents, seule la première valeur est retenue (GUID).

📌 Suivi et traçabilité

Tous les messages d'erreur et de traitement sont enregistrés pour assurer un suivi précis.

📌 Résultat

Chaque SKU est désormais unique et les données sont cohérentes pour l'analyse.

⚠️ ATTENTION! ⚠️

Les valeurs de **TOTAL_SALES différent** ici ont été sommées, parce que considérées comme des ventes cumulatives provenant de différents posts pour le même produit. Il est important de vérifier la logique métier, parce que ça pourrait fausser les résultats.

TRAITEMENT DES VALEURS NÉGATIVES

1. Problème:

- Certaines valeurs dans les colonnes PRICE, STOCK_QUANTITY et TOTAL_SALES étaient négatives, ce qui est incohérent et peut indiquer une erreur de saisie ou de système.

2. Solution:

- Conversion de toutes les valeurs négatives en NaN et ajout d'une colonne VAL_NEG pour signaler ces anomalies et permettre un suivi ou une correction ultérieure.

3. Ajout d'un indicateur des valeurs manquantes

- Création d'une colonne VAL_VIDE pour identifier les lignes contenant des valeurs vides.
- Ajout d'une colonne err_count comptant le nombre de champs vides par ligne pour un suivi futur.

Résultat:

Données nettoyées, sans incohérences, et avec un suivi des valeurs modifiées pour une meilleure traçabilité.



FUSION DES TABLES *ERP*, *WEB* ET *LIAISON*

Fusion ERP & Liaison : Aucun problème

 Les deux tables possèdent la clé `PRODUCT_ID`, ce qui permet une fusion directe sans perte de données.

Fusion ERP-Liaison & Web : Problèmes rencontrés

 Problème 1: Tous les produits ne sont pas vendus en ligne, donc certains n'ont pas de clé Web (*SKU*).

 Problème 2: La table Web contient 3 produits sans *SKU*, ce qui empêche une correspondance immédiate.

 Problème 3: Nous avons créé 2 *SKU* (*noSKU_01* et *noSKU_02*) qui ne sont pas présentes dans ERP-Liaison.

Solution : Jointure externe (*how='outer'*)

 Permet d'inclure toutes les lignes, même sans correspondance immédiate.

 Ajout de *indicator=True*: Création de la colonne `_MERGE` pour suivre les lignes sans clé correspondante.

Résultat:

- Aucune perte de données
- Identification facile des incohérences
- Correction des valeurs manquantes possible après la fusion

INCOHÉRENCES ENTRE *SKU* ET *PRODUCT_ID*

Processus d'identification:

- **2** produits WEB sans SKU ont été détectés avant la fusion (noSKU_01 et noSKU_02).
- Après l'intégration des données de l'ERP et de Liaison, **3** lignes sans SKU mais avec des attributs WEB complets et cohérents ont été trouvées.
- Grâce à ces attributs (POST_TITLE, PRICE, ONSALE_WEB, etc.), nous avons pu associer ces lignes aux bons product_id.

Correction appliquée:

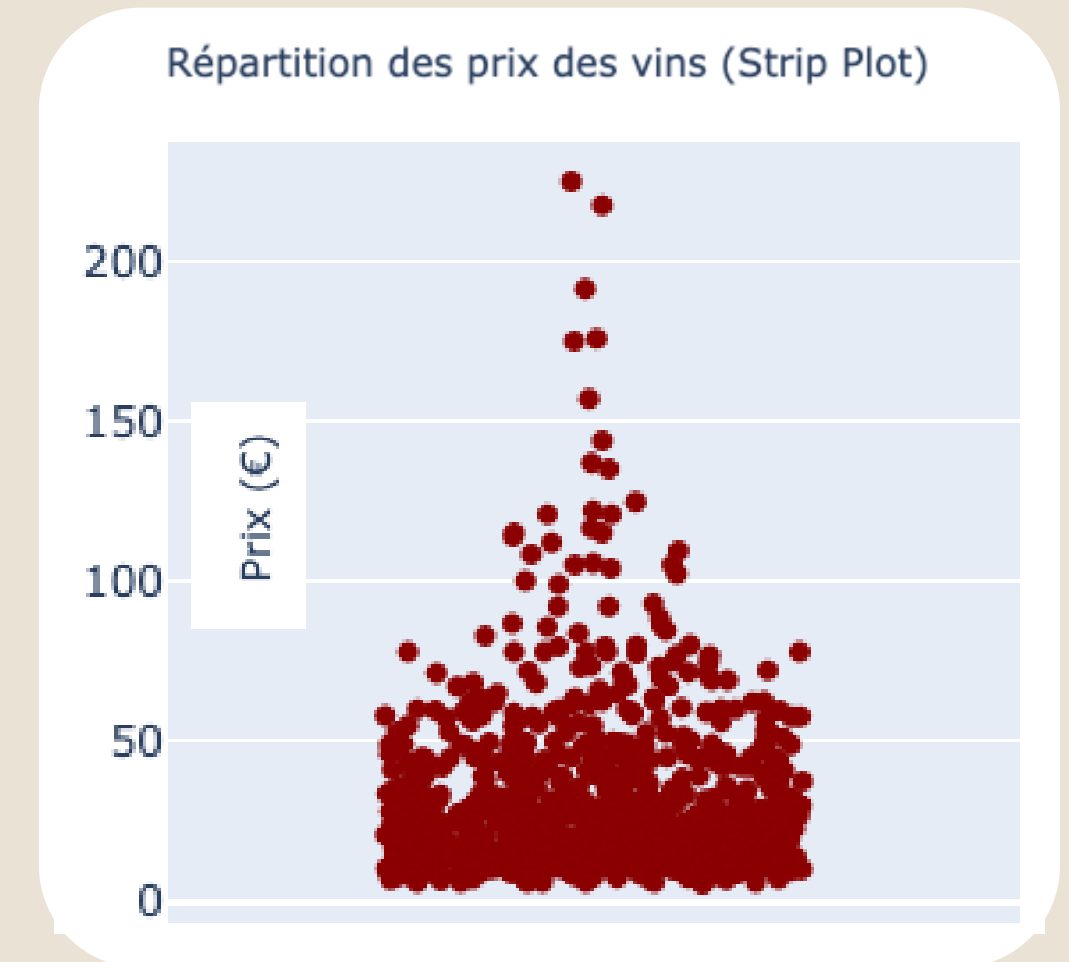
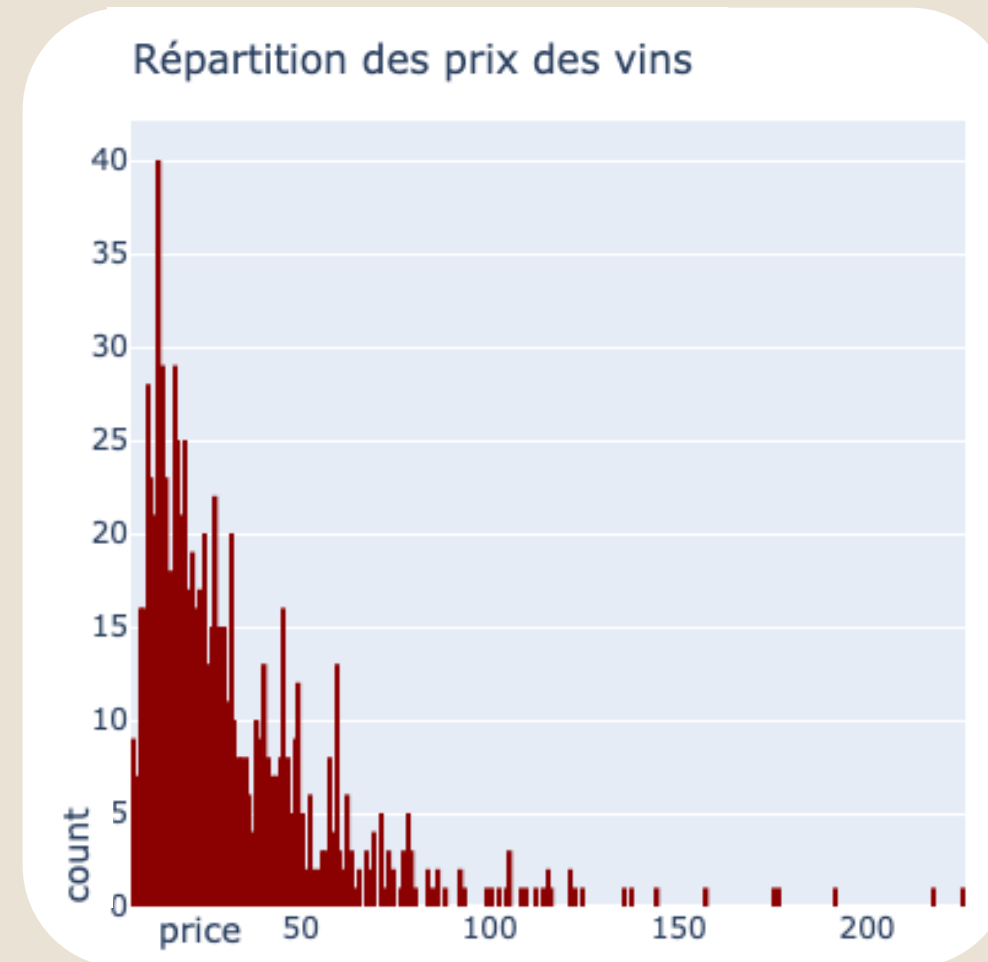
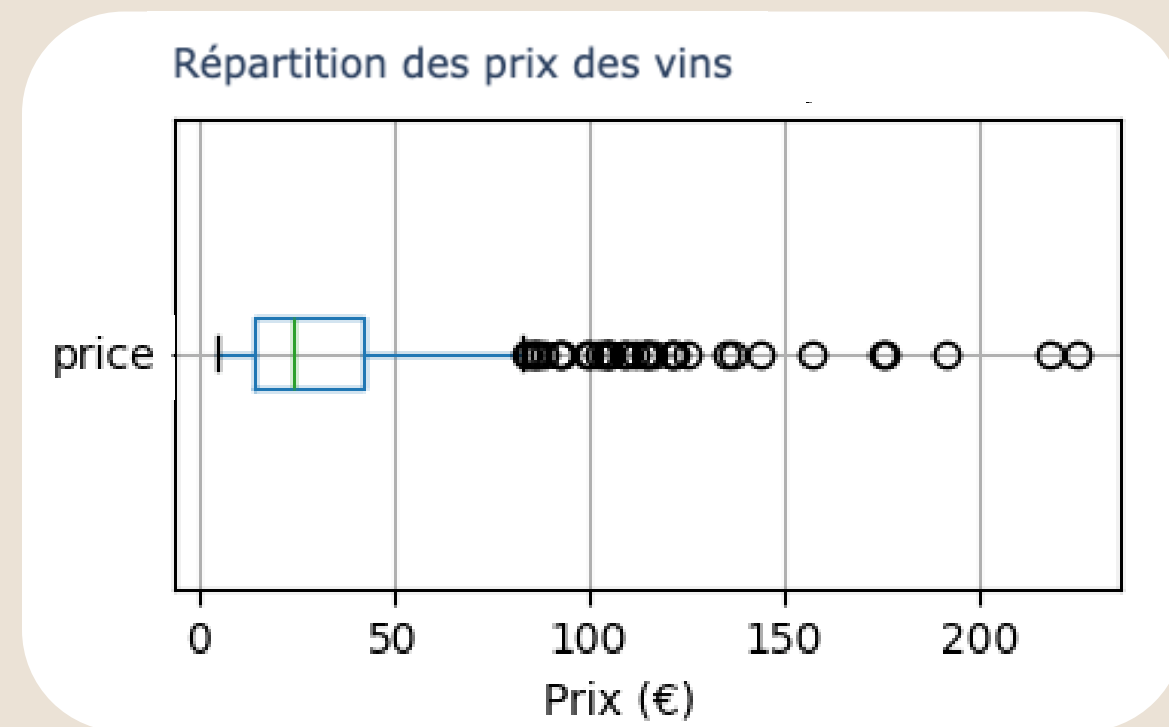
- ✓ Attribution de nouveaux SKU (noSKU_01 et noSKU_02).
- ✓ Association des bons PRODUCT_ID en croisant les informations de prix et de disponibilité.

Résultat:

Les produits ont retrouvé un SKU unique et les bonnes informations associées.
Les incohérences ont été corrigées pour garantir un catalogue fiable et exploitable.

Analyses des Prix

Visualisations: Comprendre la distribution des prix, et identifier les valeurs extrêmes (*outliers*).



Les graphiques montrent que la plupart des produits se vendent dans une plage de prix standard, avec quelques exceptions notables.

Analyses des Prix

Ces outliers ne sont pas nécessairement des erreurs de saisie, mais reflètent une grande diversité de gammes de prix, notamment dans la catégorie Whisky.

Méthodes statistiques:

IQR: 36 outliers

Z-score: 16 outliers

Prix Inhabituels: 20

Prix Extrêmes: 16

La moyenne des prix est: **32.35** euro

L'écart-type des prix est: **26.62** euro

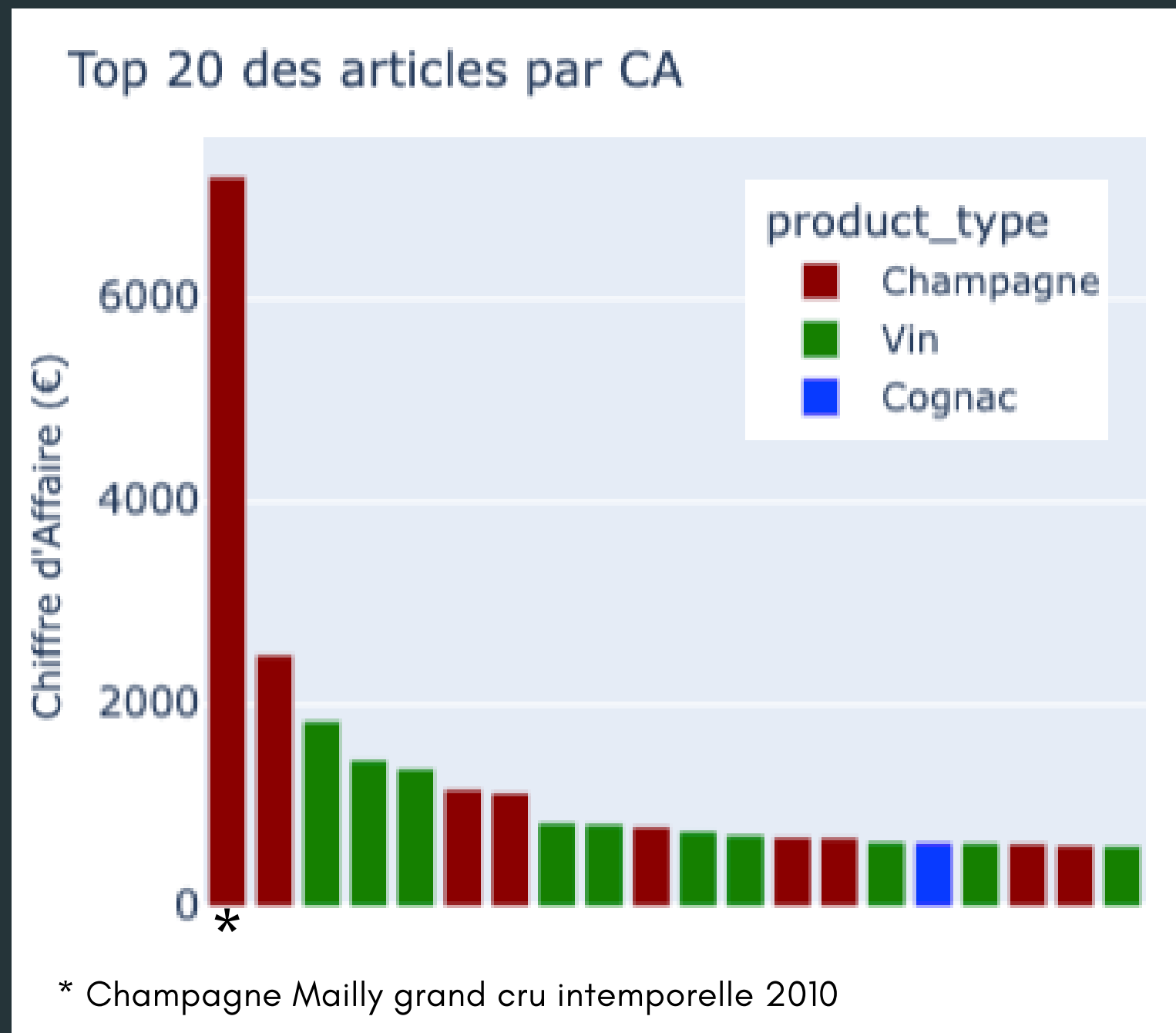
	total	%_outliers_IQR	%_outliers_Z
product_type			
Whisky	14	28.57 %	0.0 %
Champagne	28	7.14 %	3.57 %
Vin	660	4.55 %	2.27 %
Bon Cadeau	1	0.0 %	0.0 %
Cognac	8	0.0 %	0.0 %
Gin	2	0.0 %	0.0 %
Huile d'olive	3	0.0 %	0.0 %

Selon ces méthodes statistiques, la proportion des outliers dans le catalogue est de: **4.36 %**

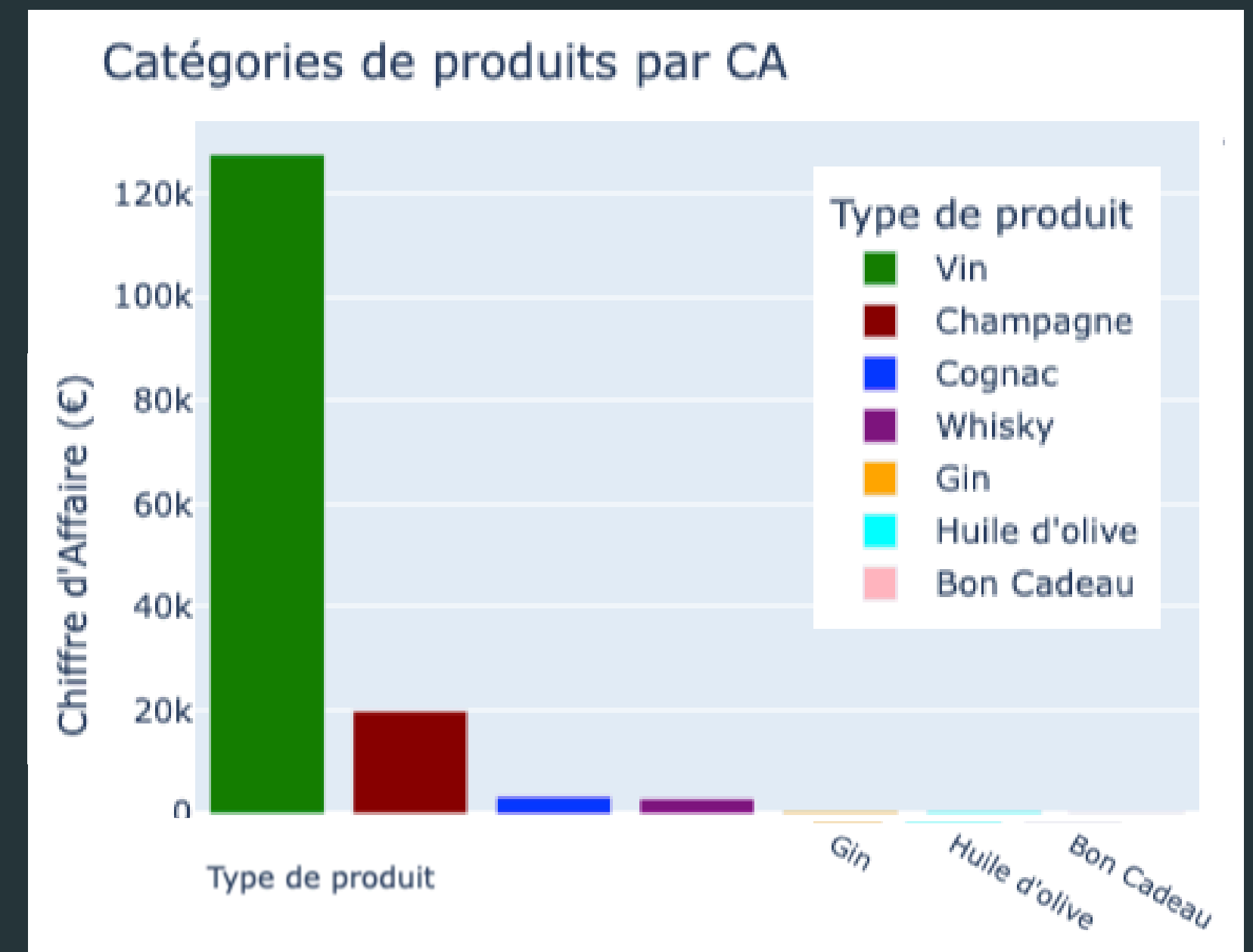
Ventes et Chiffre d'Affaires

PALMARÈS DES RÉFÉRENCES EN CA:

- ✓ Le classement des 20 articles générant le plus de CA a été établi et visualisé par un graphique en barres.



- ✓ Bien que les 20 premiers affichent principalement des Champagnes, l'analyse par catégorie révèle que, globalement, les Vins contribuent davantage au CA total.



Ventes et Chiffre d'Affaires

CALCUL DU CA:

- ✓ Le chiffre d'affaires par produit est calculé en multipliant le prix par le nombre d'unités vendues (PRICE x TOTAL_SALES).
- ✓ L'analyse se concentre sur les articles vendus sur le site web (colonne "onsale_web" = 1).

154 623 euro Chiffres d'affaires totale
(146 028 euro SANS outliers)

ANALYSE 20/80 DU CA:

L'analyse 20/80 montre la portion d'articles qui génère 80% du CA total, permettant ainsi d'identifier les références stratégiques du portefeuille.

Le nombre d'articles représentant 80% du CA est: **419**

qui correspondent à **59%** du catalogue web

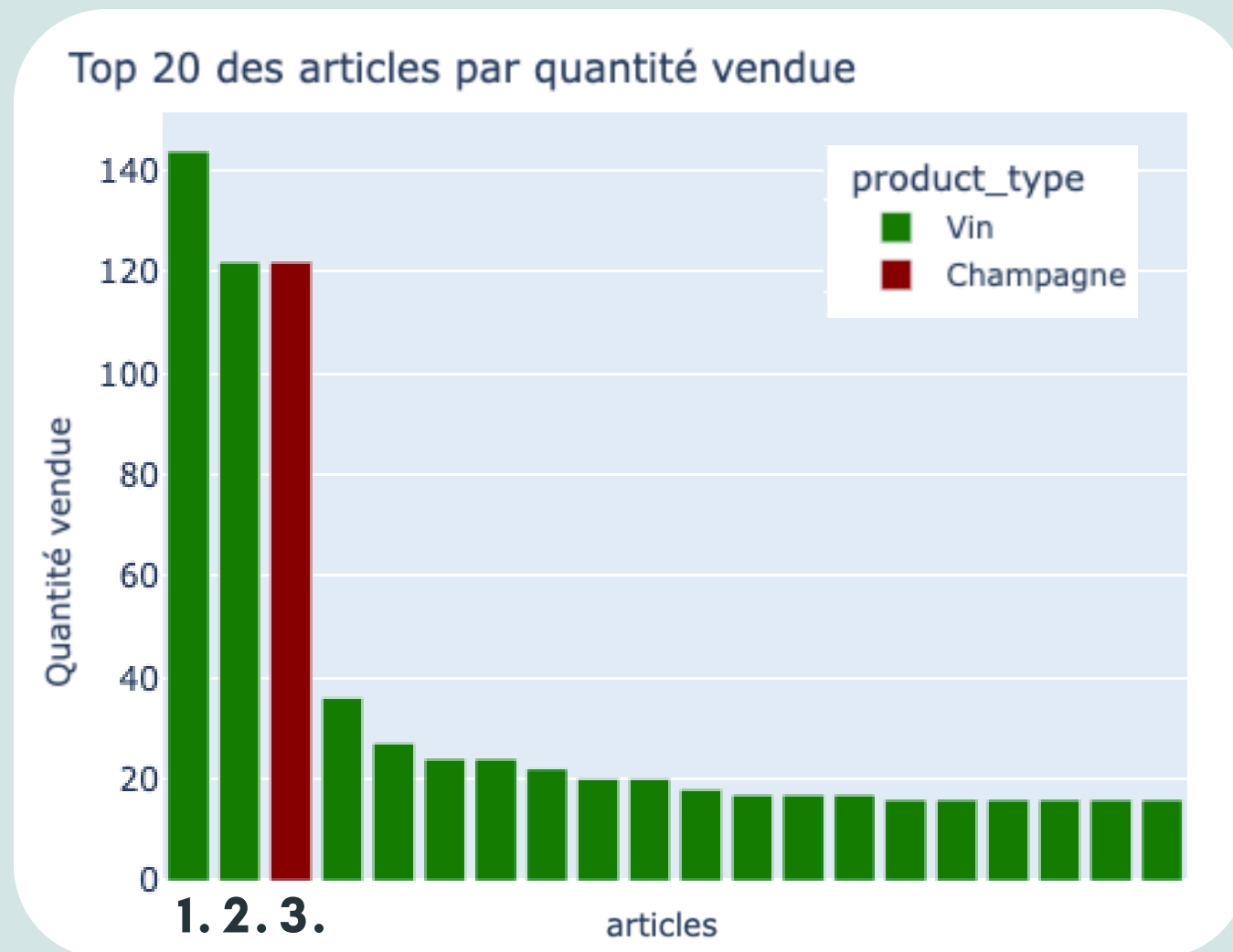


	product_type	count	percentage
0	Vin	375	89.5 %
1	Champagne	25	5.97 %
2	Whisky	8	1.91 %
3	Cognac	7	1.67 %
4	Huile d'olive	2	0.48 %
5	Gin	1	0.24 %
6	Bon Cadeau	1	0.24 %

Analyses des Quantités Vendues

Palmarès en quantités et
Analyse 20/80 sur les quantités

Le graphique présente le top 20 des articles par quantité vendue.



1. francois-baur-pinot-noir-schlittweg-2017
2. argentine-mendoza-alamos-torrentes-2017
3. champagne-mailly-grand-cru-intemporelle-2010

Le tri des produits par volume de ventes (TOTAL_SALES) révèle que la majorité des ventes est concentrée sur quelques références clés.

En calculant la part de ventes de chaque article et la somme cumulative, nous déterminons combien d'articles représentent 80% du volume total des ventes.

	product_type	count	percentage
0	Vin	407	96.22 %
1	Champagne	10	2.36 %
2	Huile d'olive	2	0.47 %
3	Cognac	2	0.47 %
4	Gin	1	0.24 %
5	Whisky	1	0.24 %

423
articles

59%
du catalogue web

Cette information est cruciale pour optimiser la gestion commerciale.

📌 Objectif:

Mesurer le nombre de mois pendant lesquels le stock actuel peut couvrir les ventes.

📌 Données utilisées:

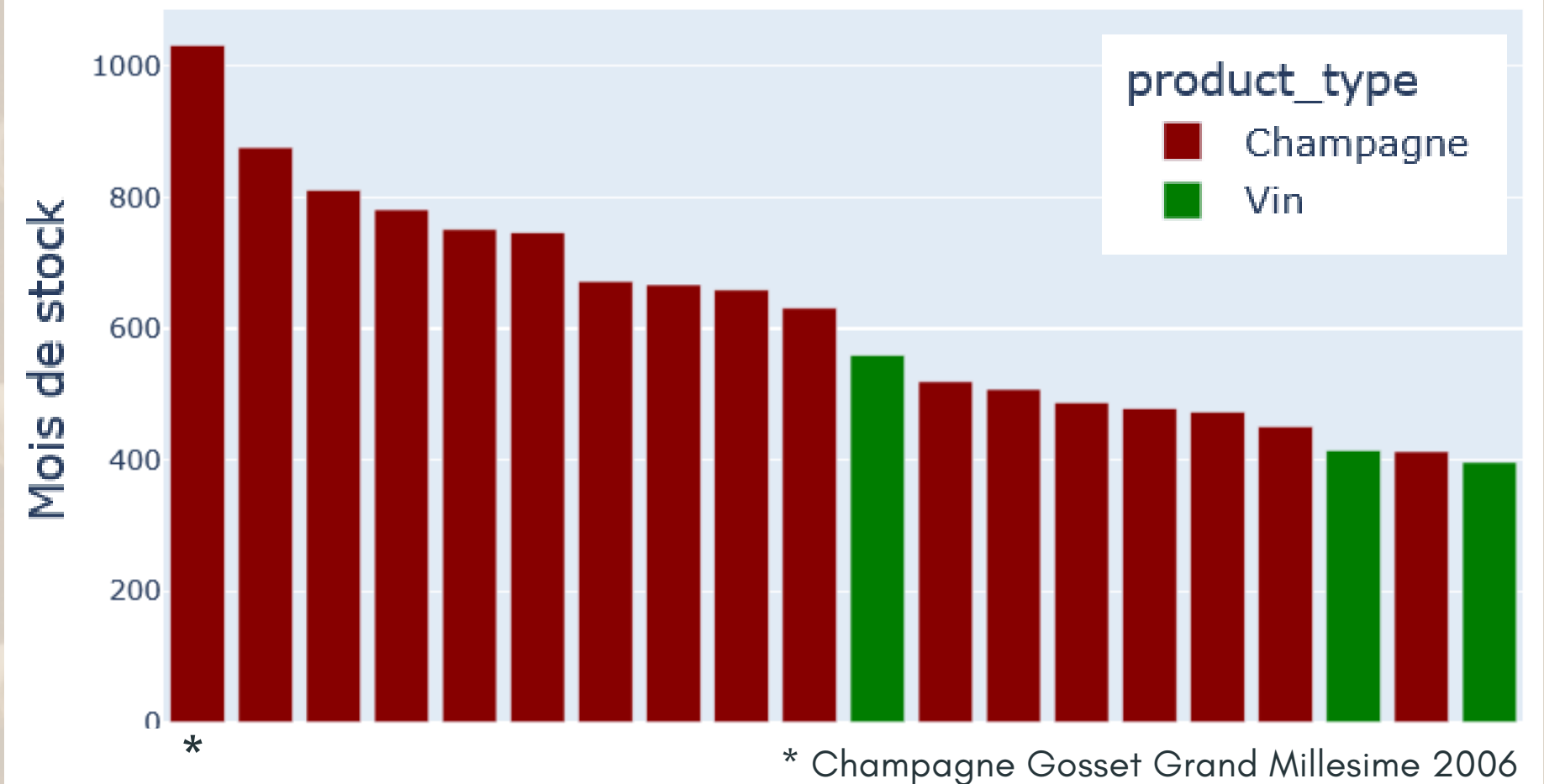
- STOCK_QUANTITY: Quantité en stock à la date d'extraction.
- TOTAL_SALES: Total des ventes réalisées.
- POST_DATE_GMT: Date de mise en vente du produit.

📌 Étapes du calcul:

1. Mois disponibles = $\text{Date d'extraction} - \text{Date de mise en vente} / 30$
2. Ventes mensuelles moyennes = $\text{total_sales} / \text{mois_disponibles}$
3. Rotation des stocks = $\text{stock_quantity} / \text{ventes mensuelles}$

LE ROTATION_STOCK INDIQUE LE NOMBRE DE MOIS DE STOCK DISPONIBLE PAR RAPPORT AU RYTHME DE VENTE, ET INTÈGRE LA DIMENSION TEMPORELLE, OFFRANT UNE ANALYSE FINE ET RÉALISTE.

Top 20 des produits avec le plus de mois de stock



Objectif:

- La valorisation des stocks est calculée pour mesurer l'investissement financier immobilisé
- Le nombre total de produits en stock est également déterminé pour évaluer la capacité d'approvisionnement.

Calcul:

Le calcul est effectué en multipliant le nombre d'unités en stock par le coût d'achat unitaire.

(STOCK_QUANTITY x PURCHASE_PRICE)

LA VALORISATION TOTALE DES STOCKS EN EUROS EST :

298 628 EUROS

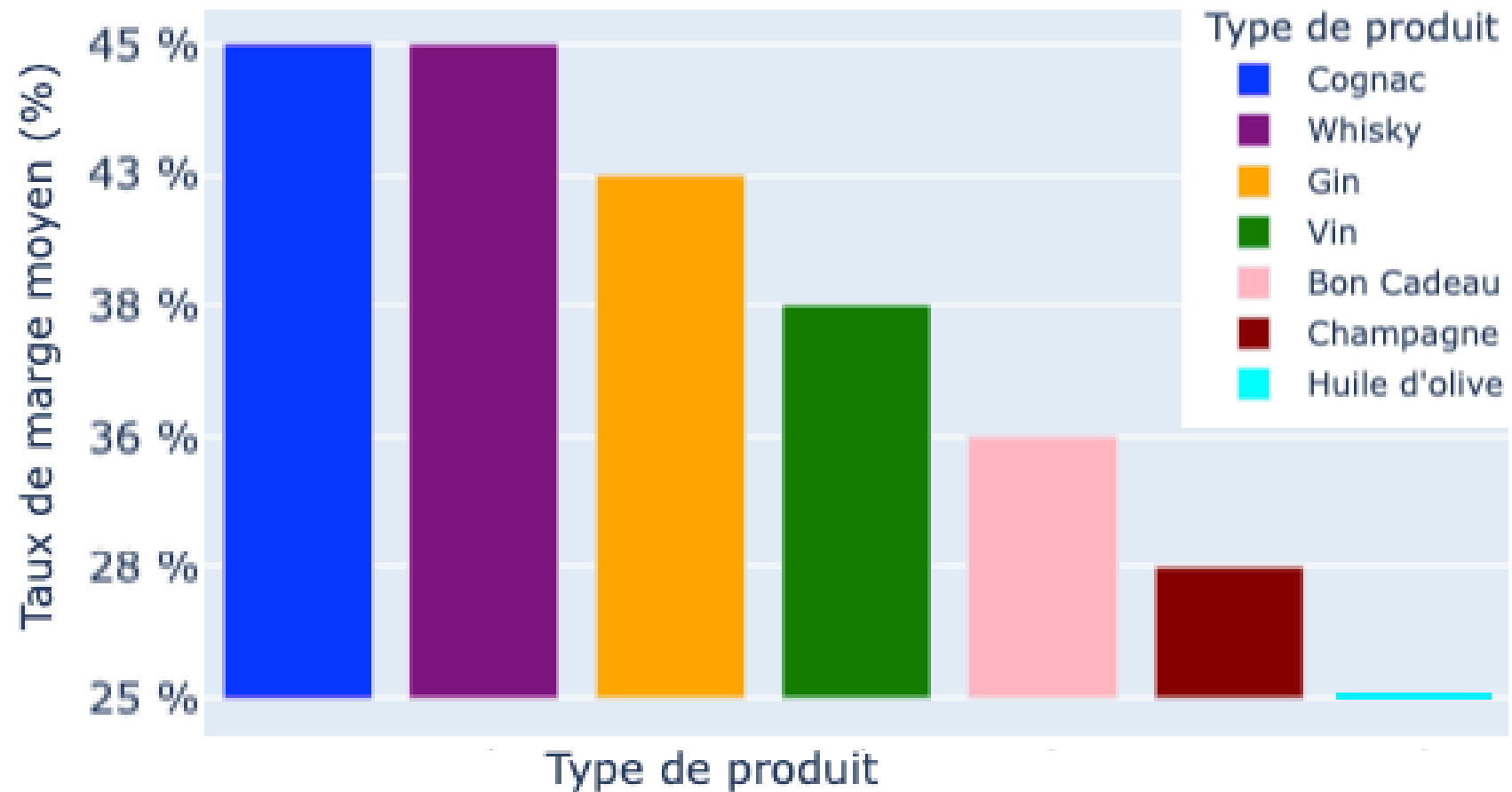
TOTAL DE PRODUITS EN STOCK :

17 822

(La somme du totale de STOCK_QUANTITY)

Analyse du Taux de Marge

Taux de marge moyen par type de produit



RÉSULTATS:

Le graphique présente le taux de marge moyen par type de produit, permettant d'identifier les segments les plus rentables et ceux nécessitant des ajustements de pricing.

Les analyses mettent en évidence des cas de taux de marge négatif et des variations de rentabilité selon les catégories de produits.

CALCUL DU TAUX DE MARGE

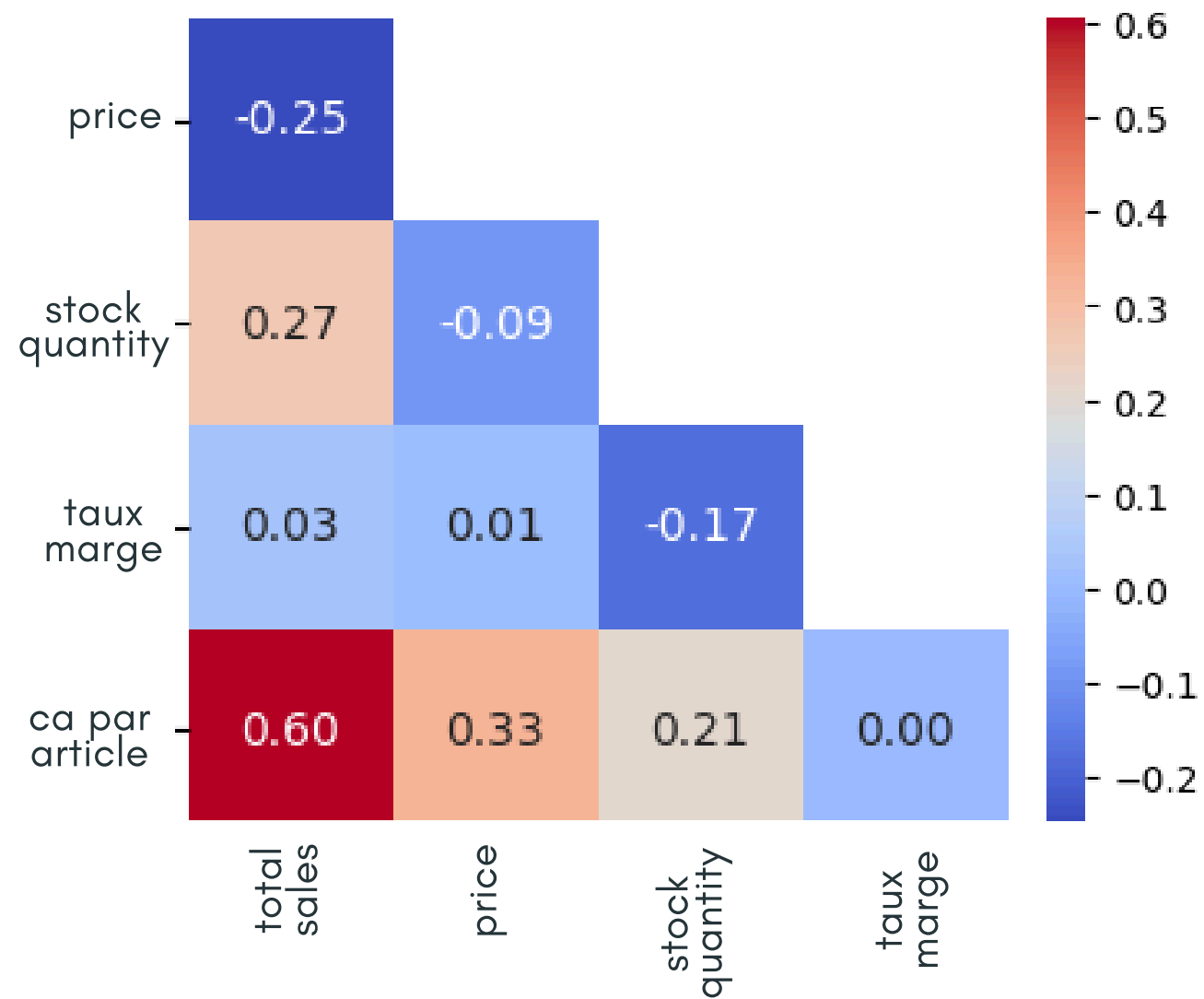
Le prix hors taxes (PRIX_HT) est obtenu en divisant le prix TTC par 1,20 (pour une TVA de 20 %).

Le taux de marge est ensuite calculé selon la formule:

$$\text{Taux de Marge} = \frac{\text{Prix HT} - \text{Coût d'Achat}}{\text{Prix HT}} \times 100$$

product_id	onsale_web	taux_marge	price	prix_HT	purchase_price
4355.0	1.0	-634.988142	12.65	10.541667	77.48
4864.0	0.0	-44.433735	8.30	6.916667	9.99
6324.0	0.0	-29.130435	92.00	76.666667	99.00
7196.0	0.0	-20.774194	31.00	25.833333	31.20

HEATMAP DES CORRÉLATIONS ENTRE VARIABLES CLÉS



CONCLUSION SUR LES CORRÉLATIONS:

L'information la plus importante est qu'un volume de ventes élevé augmente le CA et que des prix trop élevés peuvent réduire le nombre de ventes.

Ces informations sont cruciales pour décider comment fixer les prix et orienter la stratégie commerciale de l'entreprise.

Analyse des Corrélations

INTERPRÉTATION DES CORRÉLATIONS

PRICE ET CA_PAR_ARTICLE (0,33):

Bien qu'un prix plus élevé puisse contribuer à un chiffre d'affaires plus important par article, le facteur le plus décisif reste le nombre d'unités vendues.

PRICE ET TOTAL_SALES (0,25):

En augmentant les prix, le volume des ventes tend à diminuer. C'est logique, car des prix trop élevés peuvent décourager les achats.

TAUX_MARGE:

Les autres chiffres, tels que TAUX_MARGE, ne changent pas beaucoup et ne nous aident pas à mieux comprendre les ventes, ils sont donc moins importants dans cette analyse.

CA_PAR_ARTICLE ET TOTAL_SALES (0,60):

Cette corrélation est la plus utile.

Elle signifie que les produits vendus en grandes quantités génèrent naturellement plus de chiffre d'affaires.

En d'autres termes, c'est le volume des ventes qui détermine principalement le chiffre d'affaires.

Recommandations

OPTIMISER LA GESTION DES DONNÉES ET DES STOCKS
POUR UNE VISION LOINTAINE

■ QUALITÉ DES DONNÉES ET FIABILITÉ DES INDICATEURS

Assurer une base de données sans erreurs de saisie ni incohérences est essentiel pour des décisions stratégiques éclairées.

■ PERFORMANCE COMMERCIALE ET CONCENTRATION DES VENTES

Identifier les références clés (analyse 20/80) afin d'optimiser le portefeuille produits et d'orienter la stratégie de réassort.

■ INTERCONNEXIONS ENTRE VARIABLES

Utiliser les corrélations pour orienter les décisions relatives à la fixation des prix et à la gestion des volumes de vente.

■ GESTION DES STOCKS ET VALORISATION FINANCIÈRE

Comprendre la rotation des stocks et la valorisation des investissements afin d'éviter surstockage ou ruptures.

■ RENTABILITÉ

Ajuster les marges et la stratégie de pricing en fonction des performances pour maximiser la rentabilité.





Recommandations en pratique

DES ACTIONS CONCRÈTES POUR
OPTIMISER PRIX, STOCKS ET RENTABILITÉ

AJUSTER LES PRIX DES PRODUITS À FORTE MARGE MAIS À FAIBLES VENTES

Réduire le prix des produits ayant un haut taux de marge mais un faible volume de ventes, en particulier le whisky, qui présente 29 % de valeurs aberrantes IQR sur la variable prix.

OPTIMISER LES PRIX EN FONCTION DU TEMPS DE STOCKAGE

Diminuer progressivement le prix des produits en stock depuis longtemps et ayant une faible rotation, tout en maintenant une marge positive (prix supérieur au coût d'achat et aux frais).

AUGMENTER LE PRIX DES BEST-SELLERS

Revaloriser les prix des trois produits les plus vendus, car l'écart de volume de ventes avec les autres références est significatif.

Nouveau Dataset

DICTIONNAIRE

J'ai créé deux versions du dataset, l'une en français et l'autre en anglais, car il n'était pas clair dans quelle langue le système de travail était défini. Mélanger deux langues peut créer de la confusion.

Ci-dessous, je présente le dictionnaire de tous les champs du dataset que j'ai nettoyé et enrichi.

dataset FR	dataset EN	Description	Count	Type
id_produit	product_id	Identifiant unique du produit.	825	float64
sku	sku	Référence unique du produit (Stock Keeping Unit).	716	object
_merge	_merge	Indicateur d'intégration des données entre sources.	825	category
nom_produit	product_name	Nom du produit.	716	object
type_produit	product_type	Catégorie ou type de produit (ex: Vin, Champagne, Whisky).	716	object
prix	price	Prix de vente du produit TTC.	822	float64
prix_HT	net_price	Prix de vente hors taxes.	822	float64
prix_achat	purchase_price	Prix d'achat du produit.	825	float64
tax_status	tax_status	Statut fiscal du produit (ex: taxable, exonéré).	716	object
stock_status	stock_status	Indique si le produit est en stock ou en rupture.	823	object
stock_quantite	stock_quantity	Quantité actuelle en stock.	823	float64
valorisation_stock_euros	stock_valuation_euros	Valeur totale du stock en euros (quantité * prix d'achat).	823	float64
site_web	onsale_web	Indique si le produit est en vente sur le site web.	825	float64
auteur	post_author	Identifiant de l'auteur qui a publié le produit.	716	float64
mise_en_ligne	date_of_sale	Date de mise en ligne du produit sur le site.	716	datetime64[ns]
mois_disponibles	months_for_sale	Nombre de mois depuis la mise en vente du produit.	716	float64
ventes_mensuelles	monthly_sales	Nombre moyen d'unités vendues par mois.	714	float64

Nouveau Dataset

DICTIONNAIRE

dataset FR	dataset EN	Description	Count	Type
rotation_stock	rotation_stock	Ratio indiquant la vitesse de rotation des stocks.	688	object
total_ventes	total_sales	Nombre total d'unités vendues.	714	float64
ca_par_article	turnover_per_item	Chiffre d'affaires généré par article.	714	float64
taux_marge	margin_rate	Taux de marge du produit.	822	float64
outlier_IQR	outlier_IQR	Indicateur de valeur aberrante selon l'IQR.	825	bool
outlier_Z	outlier_Z	Indicateur de valeur aberrante selon le score Z.	825	bool
VAL_VIDE	VAL_VIDE	Indicateur de valeurs manquantes.	825	bool
VAL_NEG	VAL_NEG	Indicateur de valeurs négatives suspectes.	825	bool
VAL_NOM	VAL_NAME	Indicateur des produits qui semblent double dans le nom.	825	bool
err_count	err_count	Nombre des champs vides à remplir \ corriger.	825	int64
z_score	z_score	Score Z du prix, sur le totale.	822	float64
z_score_type	z_score_type	Score Z du prix, par type de produit.	713	float64
prix_moyen_type	mean_price_type	Prix moyen, par type de produit.	716	float64
prix_mediane_type	median_price_type	Prix médian, par type de produit.	716	float64
ecart_type_prix_type	std_price_type	écart-type du prix, par type de produit.	715	float64
Q1_prix_type	Q1_price_type	Premier quartile du prix, par type de produit.	716	float64
Q3_prix_type	Q3_price_type	Troisième quartile du prix, par type de produit.	716	float64
IQR_prix_type	IQR_price_type	Intervalle interquartile du prix, par type de produit.	716	float64

Les dimensions des NOUVEAUX DATASETS sont:

825 lignes
35 colonnes



Conclusion

Ce projet a permis de moderniser les outils existants en une analyse structurée et exploitable.

Dans le nouveau dataset, j'ai intégré plusieurs colonnes pour identifier les valeurs manquantes et ainsi vous aider à continuer à améliorer votre ERP.

Le dataset est désormais cohérent et exempt d'erreurs, et constitue un outil fiable et performant pour appuyer vos décisions stratégiques.

MERCI POUR L'ATTENTION,
ET LA CONFIANCE

Giulia Governatori