

Optimisez la gestion des données d'une boutique avec R ou Python

Rondeau Cécile
Data Analyst
3 Février 2026



Analyses Exploratoires des Données

Datasets et caractéristiques



ERP

- ▣ Lignes : 825
- ▣ Colonnes: 6

- **Product_id**
 - ▣ Onsale_web
 - ▣ Price
 - ▣ Stock_quantité
 - ▣ stock_status
 - ▣ Purchase_prix

WEB

- ▣ Lignes: 1513
- ▣ Colonnes: 29

- **Sku**
 - ▣ Total_sales
 - ▣ Post_date
 - ▣ Product_type
 - ▣ Post_title

● ▣ Autres colonnes: 24
métadonnées techniques
et éditoriales

Liaison

- ▣ Lignes: 825
- ▣ Colonnes: 2

- **Product_id**
 - **Id_Web**

Traitements réalisés - Nettoyage des données caractéristiques -ERP



	product_id	onsale_web	price	stock_quantity	stock_status	pi
	4	4039	1	46.0	3	outofstock
398	4885	1	18.7	0	instock	

Quantité Stock
en négatif

Incohérence
entre statut et
la quantité du
stock

Prix négatif

Extraction des
enregistrements
incohérents
df_erp_incoherent

	product_id	onsale_web	price	stock_quantity	stock_status
449	4973	0	10.0	-10	outofstock
573	5700	1	44.5	-1	outofstock

	product_id	onsale_web	price
151	4233	0	-20.0
469	5017	0	-8.0
739	6594	0	-9.1

Création d'un
dataset nettoyé
df_erp_clean

Lignes: 818
Colonnes: 6

Traitements réalisés - Nettoyage des données caractéristiques - WEB

715 Skus apparaissent au moins une fois en doublon dans le jeu de données.

	sku	total_sales	post_date	product_type	post_title
668	10014	10.0	2019-04-04 15:45:23	Gin	Damley's London Dry Gin Original
1030	10014	10.0	2019-04-04 15:45:23	Gin	Damley's London Dry Gin Original
748	10459	4.0	2018-04-13 15:58:19	Vin	Alphonse Mellot Sancerre Rouge Génération XIX ...
887	10459	4.0	2018-04-13 15:58:19	Vin	Alphonse Mellot Sancerre Rouge Génération XIX ...
1317	10776	6.0	2018-04-17 21:28:52	Vin	Albert Mann Pinot Gris Vendanges Tardives Alle...



	sku	total_sales	
1084	NaN	-56.0	2
1087	NaN	-17.0	1

Sku données non numérique

Sku non uniques

Lignes vides

Montants négatifs

Extraction des enregistrements incohérents
df_web_incoherent

Suppression des 24 colonnes

Création d'un dataset nettoyé
df_web_clean

Lignes: 712
Colonnes: 5

Nombre de SKU non numériques (str) : 4

sku	
272	13127-1
842	bon-cadeau-25-euros
1117	13127-1
1387	bon-cadeau-25-euros

Nombre de lignes sans SKU : 85

	sku	total_sales	post_date	product_type	post_title
639	NaN	NaN	NaT	NaN	NaN
199	NaN	NaN	NaT	NaN	NaN
1244	NaN	NaN	NaT	NaN	NaN
1429	NaN	NaN	NaT	NaN	NaN

Traitements réalisés - Nettoyage des données caractéristiques - Liaison



Nombre de id_web non numériques (str) : 3	
[17]:	id_web
443	bon-cadeau-25-euros
822	13127-1
823	14880-1



Extraction des 3
identifiants Web non
conformes
incohérents
df_liaison_incoherent

df_liaison_clean

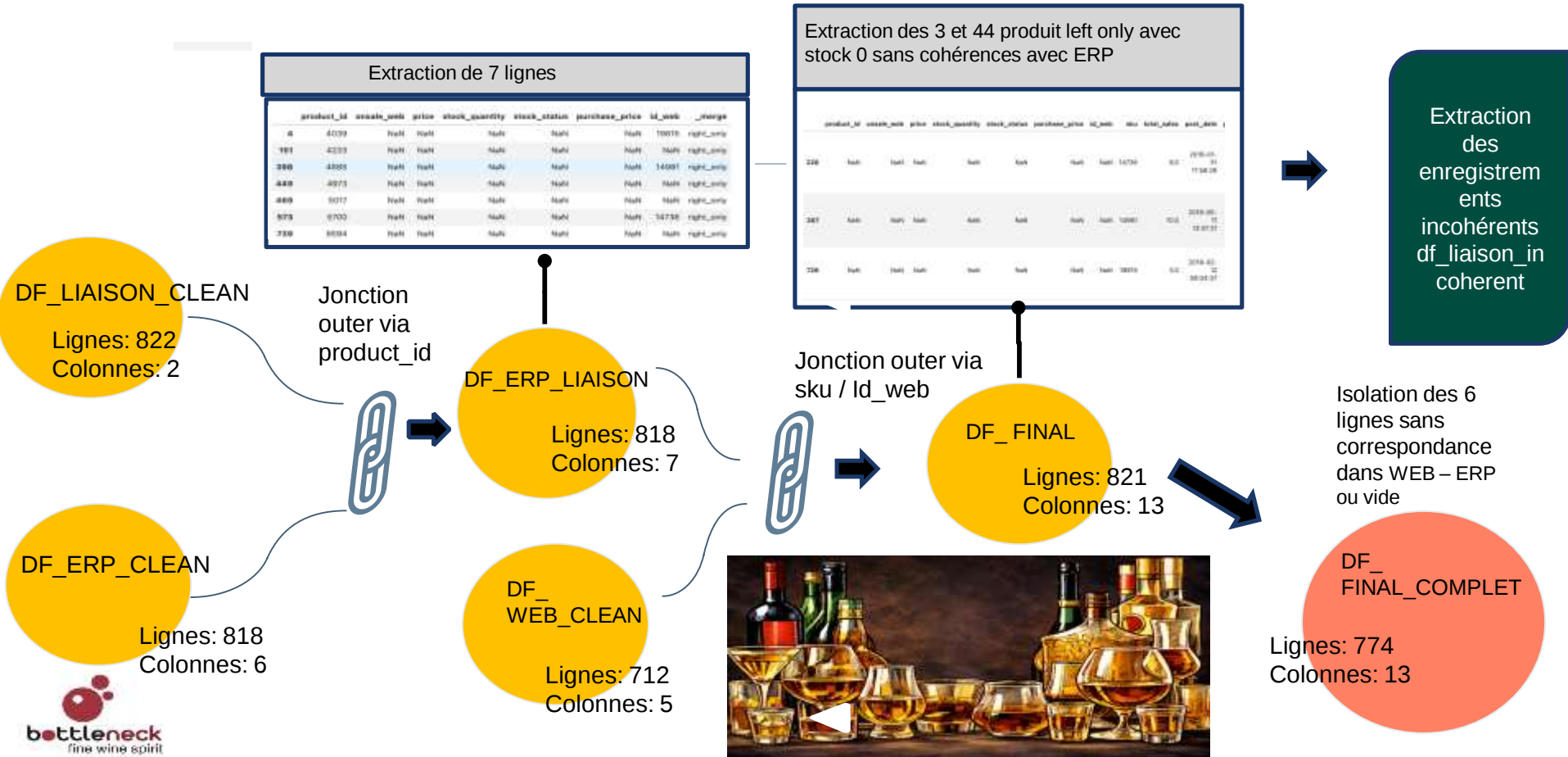
Lignes: 822
Colonnes: 2

product_id
-
Lignes 825

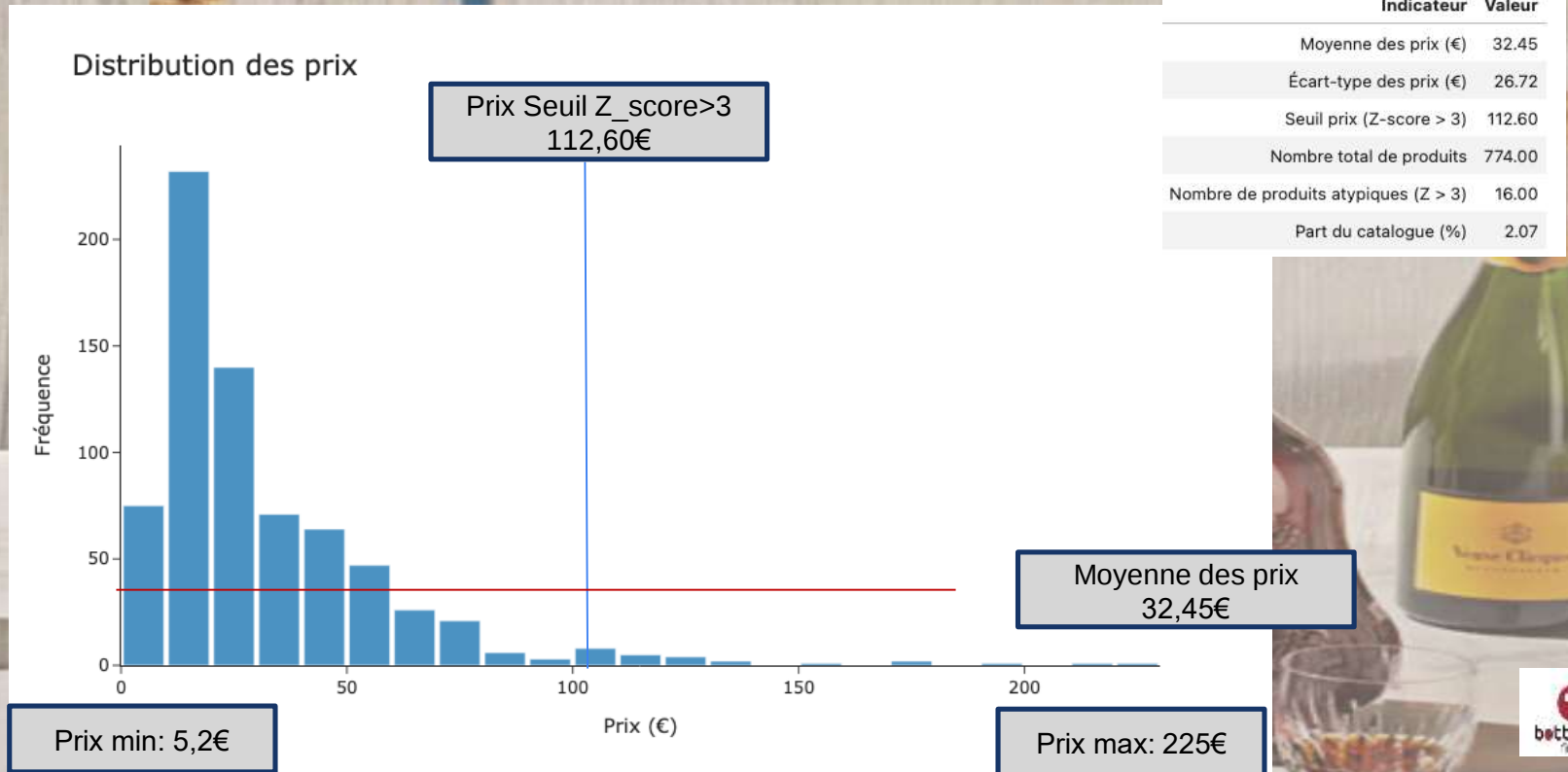
id_web
-
Lignes 734

Analyses Exploratoires des Données

Traitements réalisés - Nettoyage des données



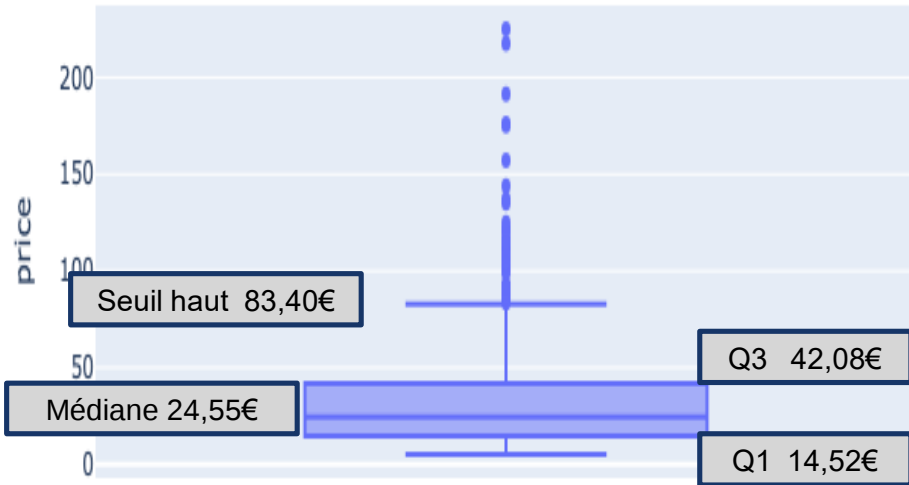
Détection des prix par la méthode du Z-score



Détection des prix atypiques par l'intervalles interquartile (IQR)

Nombre d'articles outliers : 33
Nombre total d'articles : 774
Proportion d'outliers dans le catalogue : 4.26%

Répartition des prix



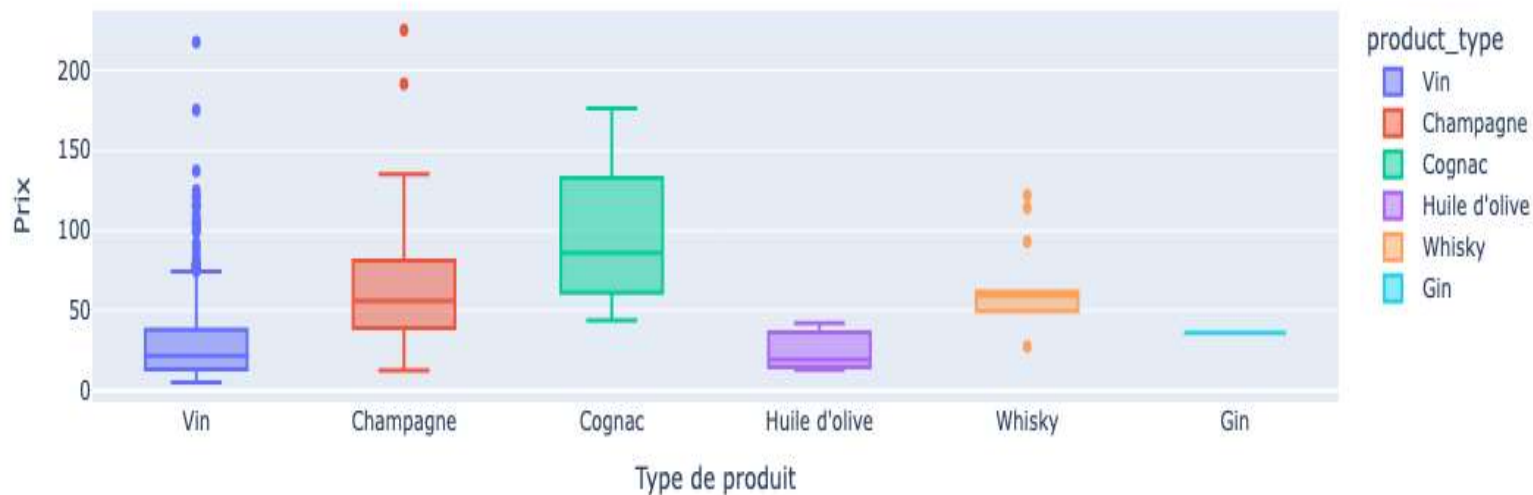
Indicateur	Valeur (€)
Premier quartile (Q1)	14.52
Médiane (Q2)	24.55
Troisième quartile (Q3)	42.08
Intervalle interquartile (IQR)	27.55
Seuil haut ($Q3 + 1,5 \times IQR$)	83.40



Analyse des prix par type de produit



Distribution des prix par type de produit

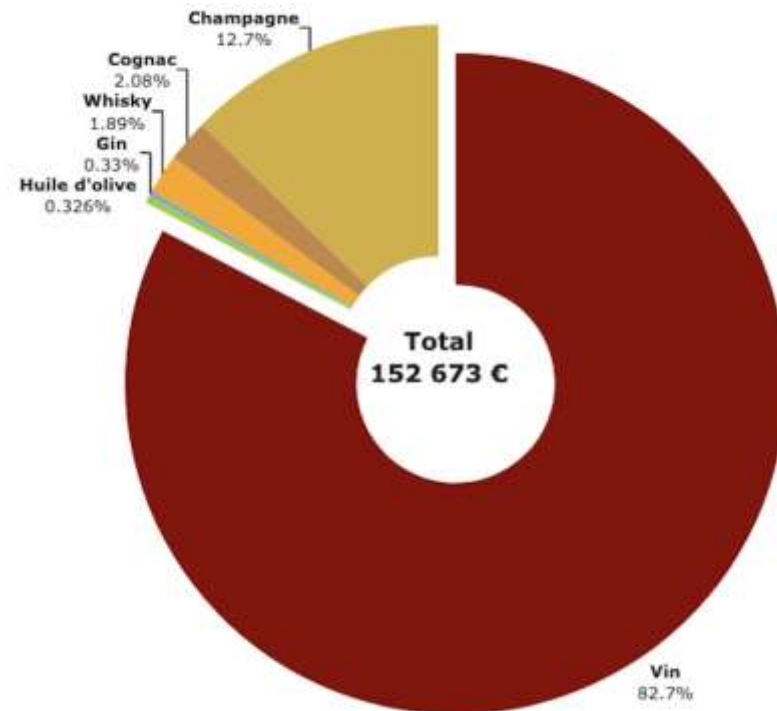


Chiffre d'affaires et Répartition

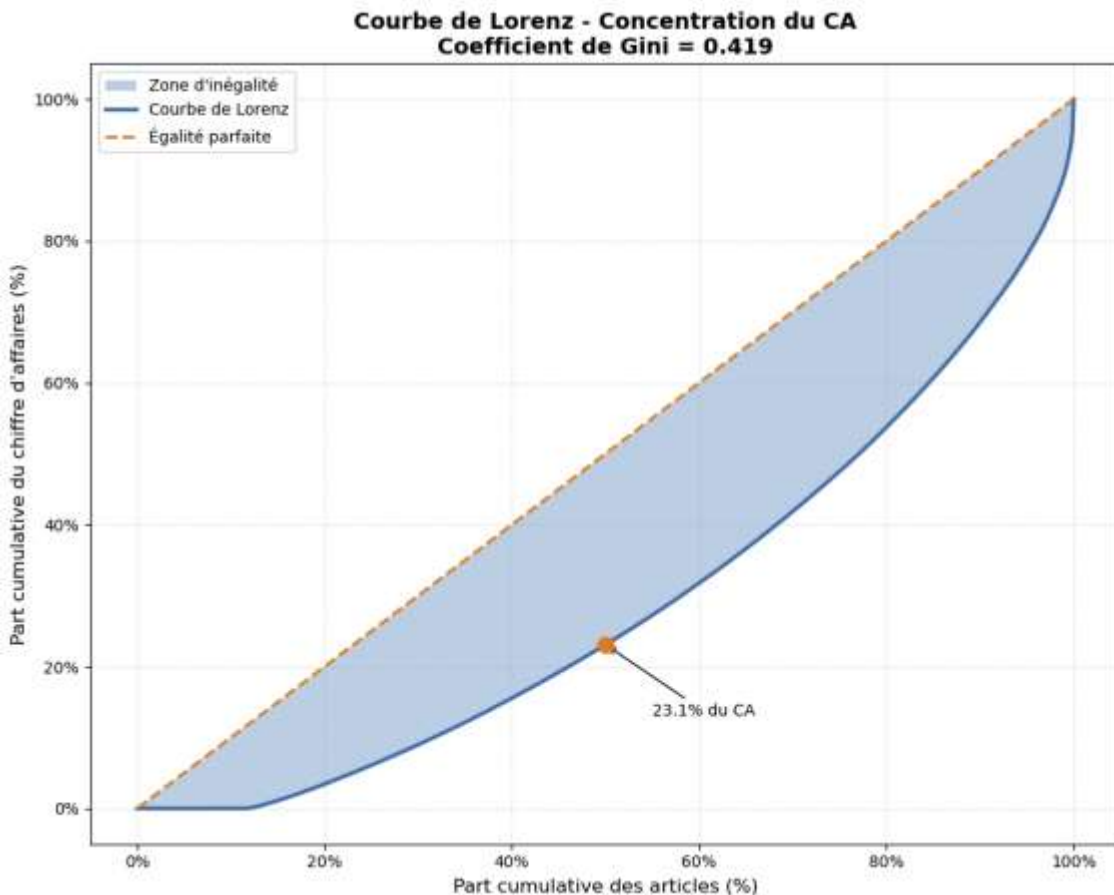


RÉPARTITION DU CHIFFRE D'AFFAIRES

Total: 152 673 €



Concentration du chiffre d'affaires – courbe de Lorenz - Gini



ANALYSE DE CONCENTRATION DU CHIFFRE D'AFFAIRES

Nombre d'articles analysés: 774
Coefficient de Gini: 0.419

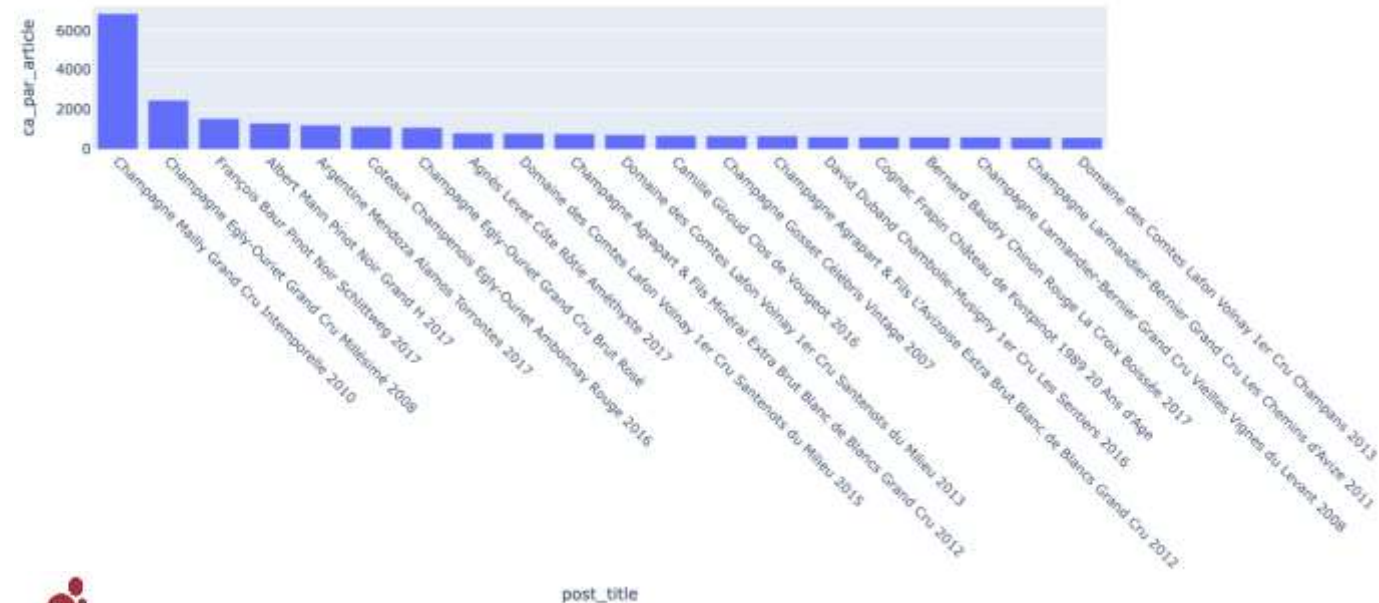
Les 20% articles les moins rentables génèrent 3.4% du CA
→ Les 80% articles les plus rentables génèrent 96.6% du CA

Les 50% articles les moins rentables génèrent 23.1% du CA
→ Les 50% articles les plus rentables génèrent 76.9% du CA

Les 80% articles les moins rentables génèrent 53.7% du CA
→ Les 20% articles les plus rentables génèrent 46.3% du CA

Principaux contributeurs au chiffre d'affaires (Top 20 produits)

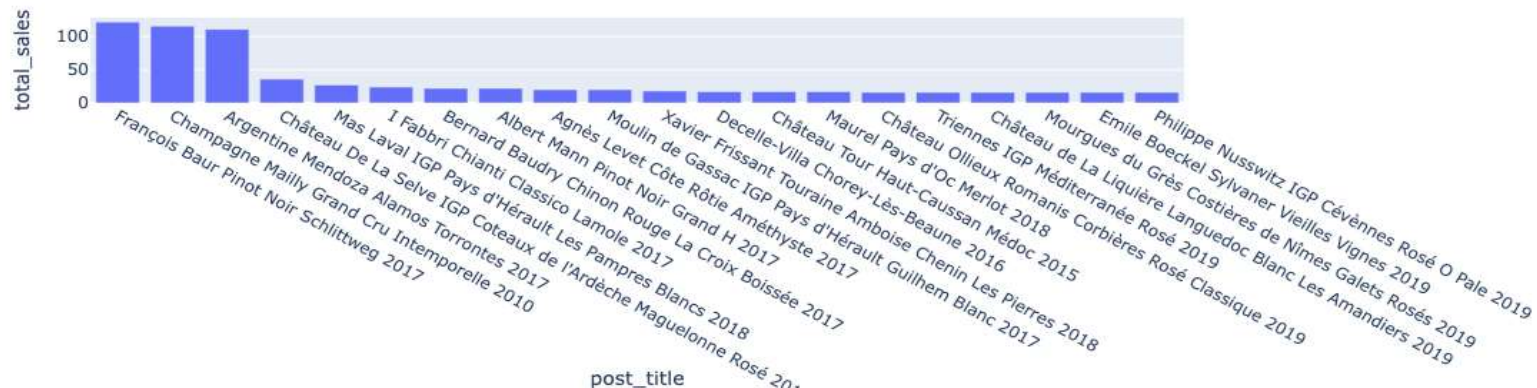
Top 20 des articles par chiffre d'affaires



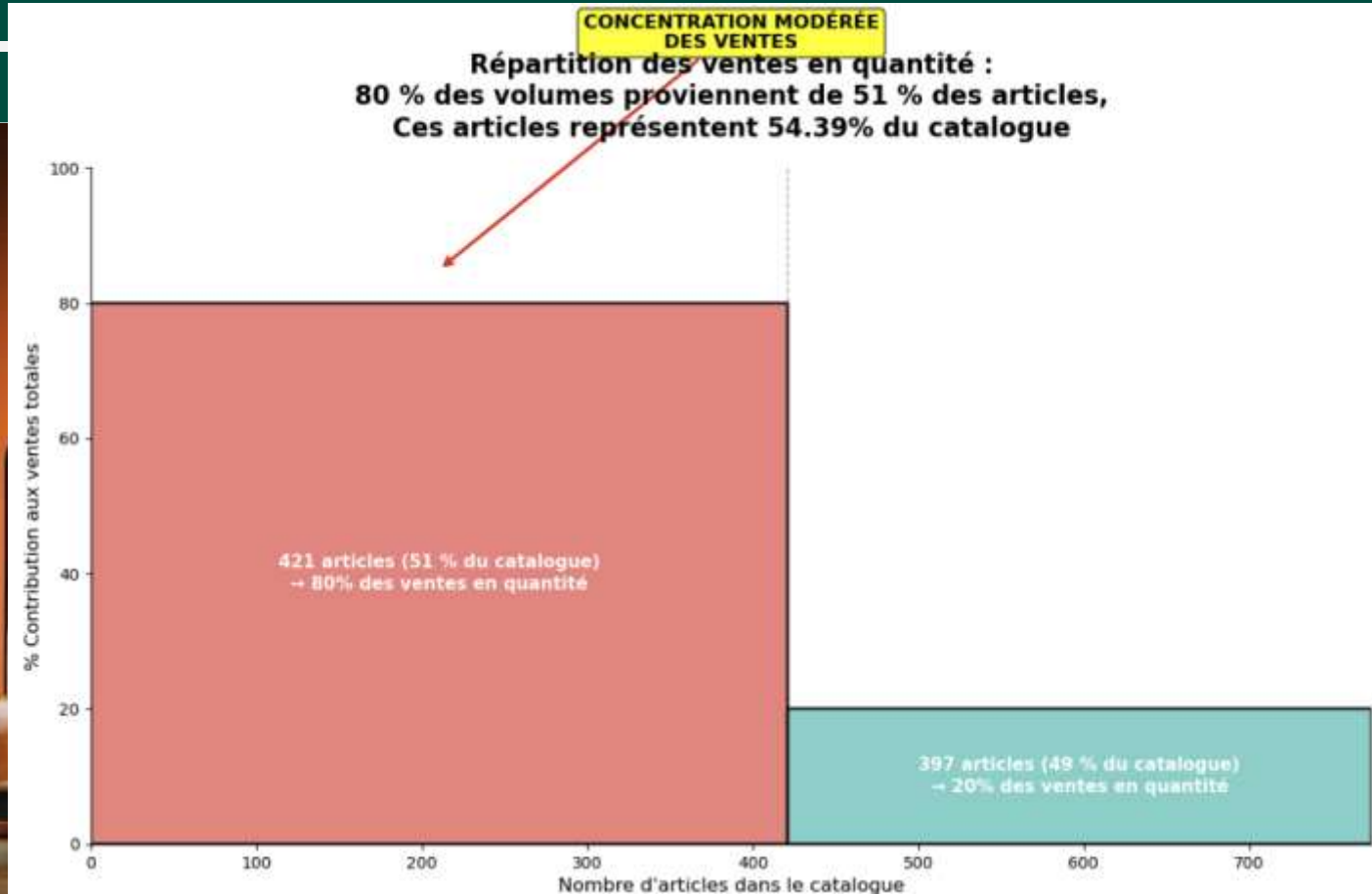
Produits les plus vendus (Top 20 en volume)



Top 20 des articles les plus vendus



Concentration des ventes- Principe 20/80

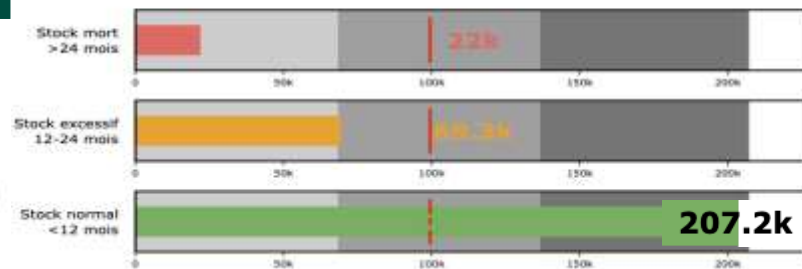
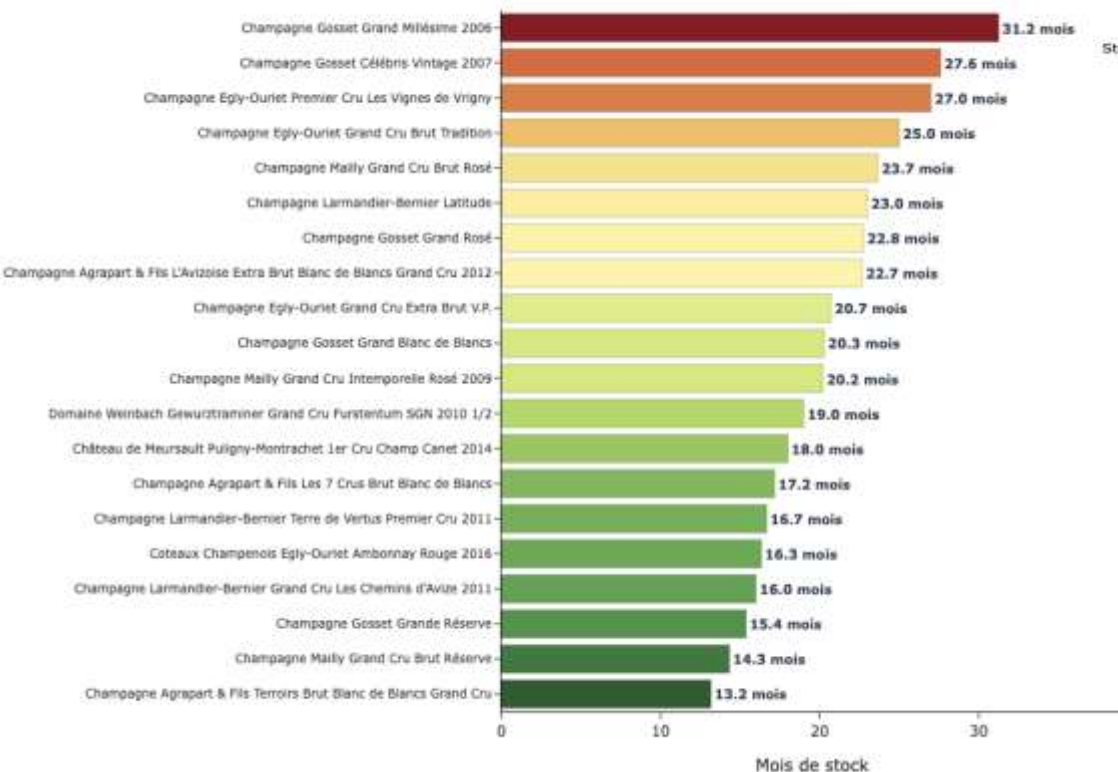


Diagnostic global des stocks

VALEUR DU STOCK PAR CATÉGORIE

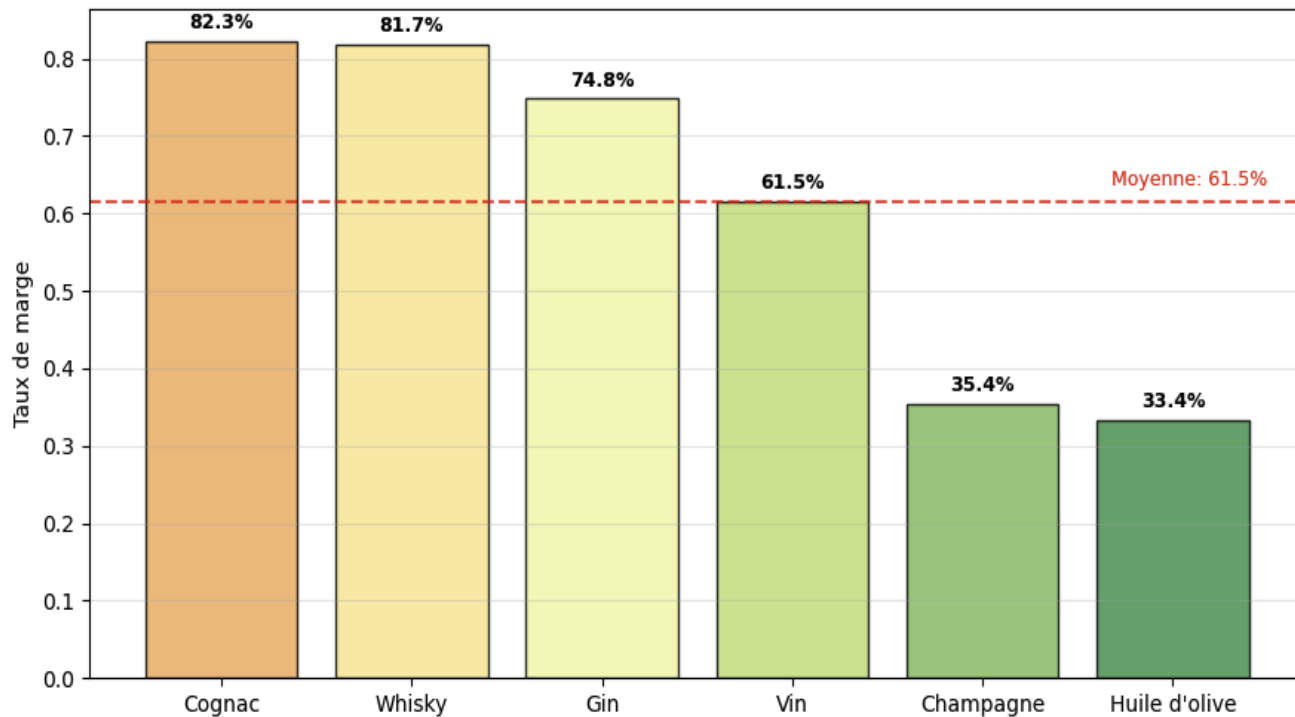
Rotation moyenne: 2.6 mois • Total: 298.5k €

Top 20 des produits avec le plus de mois de stock

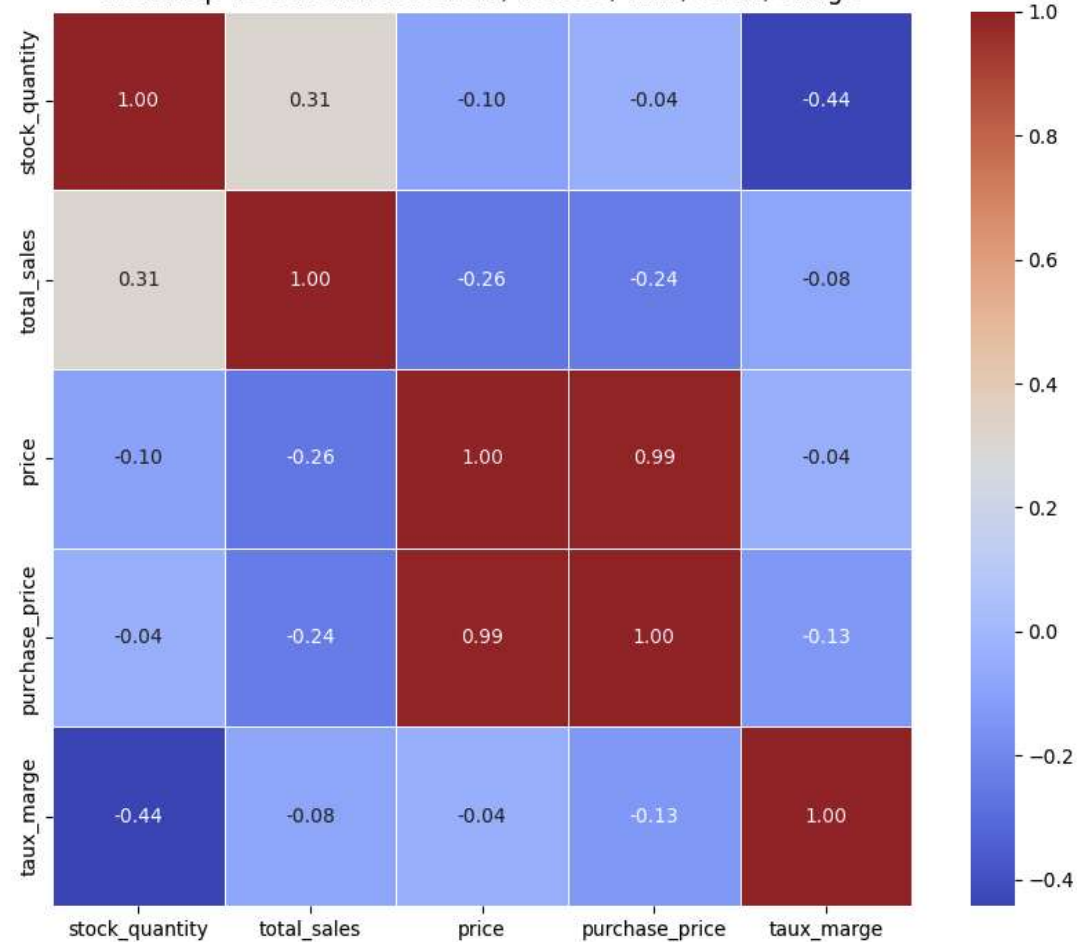


Prix et marges

Marges par Type de Produit



Heatmap de corrélation : Stock / Ventes / Prix / Coût / Marge



Optimisation de la marge: un enjeu de stock plus que de ventes



Actions pour la suite



**Améliorer la
qualité des
données en
amont**

**Sécuriser la
rentabilité
produit**

**Optimiser la
marge via les
stocks**

**Prioriser les
produits à forte
valeur**

**Adapter la
stratégie par
catégorie**

**Piloter via des
indicateurs
clés**

Points sur les compétences apprises

Ce qui s'est bien passé

Structure du nettoyage et rigueur des traitements

Capacité à repérer et comprendre les incohérences de données

Meilleure vision de la qualité globale des données

Le plus difficile

Faire le choix sur les données incohérentes à conserver, à exclure ou à modifier

Comprendre les écarts lors des jointures entre les différentes sources

Tâche à améliorer

Encore mieux développer le lien entre données et décisions métier

Mieux expliquer les résultats

Merci pour votre attention

