

Bottleneck



Nicolas Pautet
Parcours Data Analyst
Mars 2025

Contexte du projet

Analyser des données de vente et de stock

Sources de données difficiles à exploiter manuellement :

- extraction ERP
- extraction site web

sur une même période, le mois d'octobre

Objectif à terme :

avoir une procédure de traitement des données plus rigoureuse

Analyses exploratoires des données

3 Datasets (format xlsx) :

- extraction ERP : 825 lignes, 6 colonnes
- extraction site web : 1513 lignes, 29 colonnes
 - beaucoup d'informations inutiles pour l'analyse
- table de liaison : 825 lignes, 2 colonnes

Vérification et traitement de doublons et de champs vides

Vérification et correction de la cohérence des données

Quelques exemples de données problématiques

Extraction ERP :

- quantités en stock négatives
ou pas cohérentes avec le booléen de disponibilité
- prix de vente négatifs

Extraction site web :

- lignes presque sans donnée
- sku manquant
- quantités vendues négatives

Fusion et consolidation des données

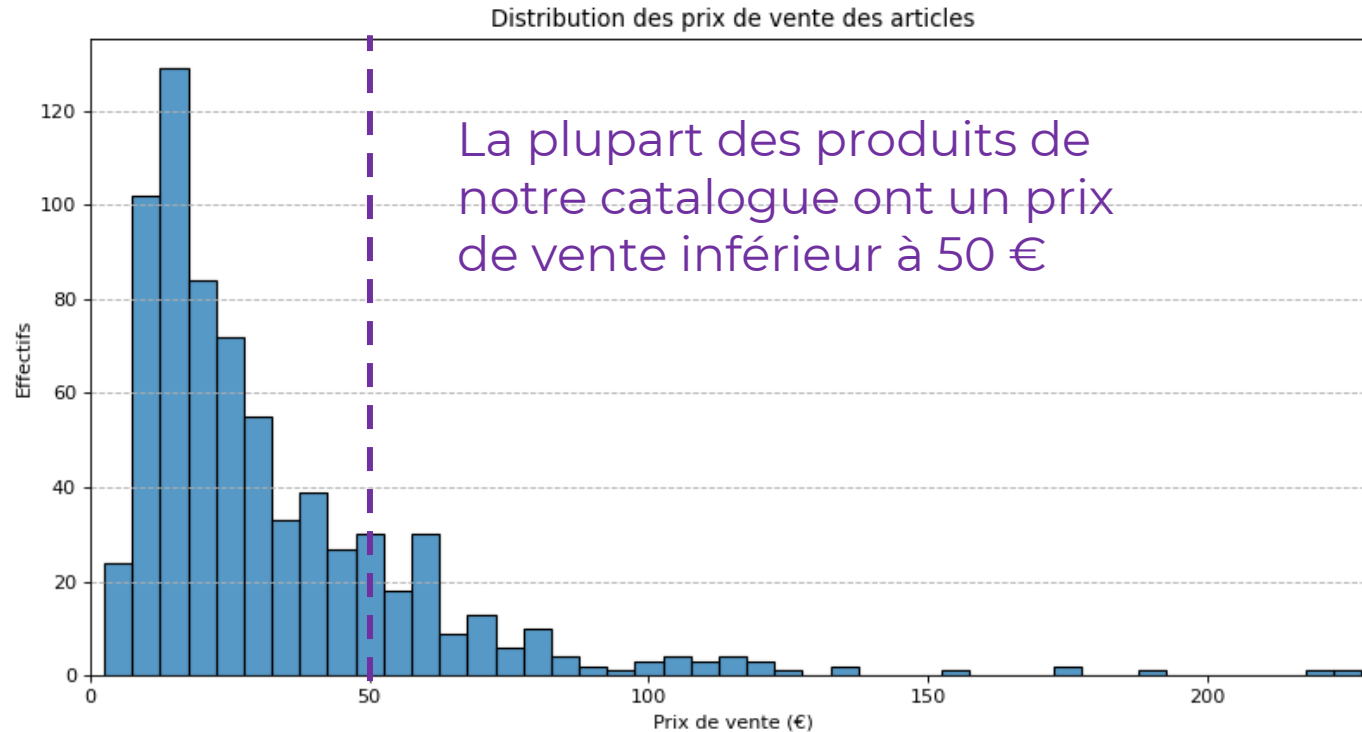
Double jointure grâce à la table de liaison,
avec pour conditions de jointure :

liaison.id_web = web.sku et liaison.product_id = erp.product_id

Jointure externe : assure la traçabilité de toutes les données,
quitte à devoir filtrer les DataFrames par la suite

Exemple : produits en stock (dans l'ERP)
mais pas disponibles pour la vente en ligne

Analyses univariées du prix



Analyses univariées du prix

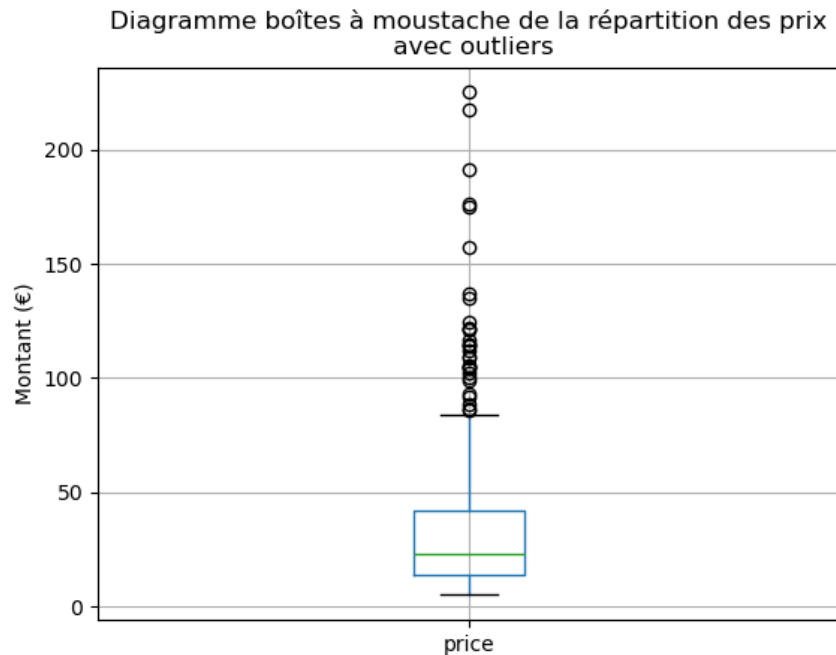
Analyse par la **méthode de l'écart interquartile** : boxplot

Q3 = 42,08 €

med = 23,45 €

Q1 = 14,06 €

Les *outliers* sont situés au-delà de **$Q3 + 1,5 * (Q3 - Q1) = 84,10$ €**



Analyses univariées du prix

Analyse par la **méthode du Z-score**

Suppose que les **prix suivent une loi normale** $N(m, \sigma)$

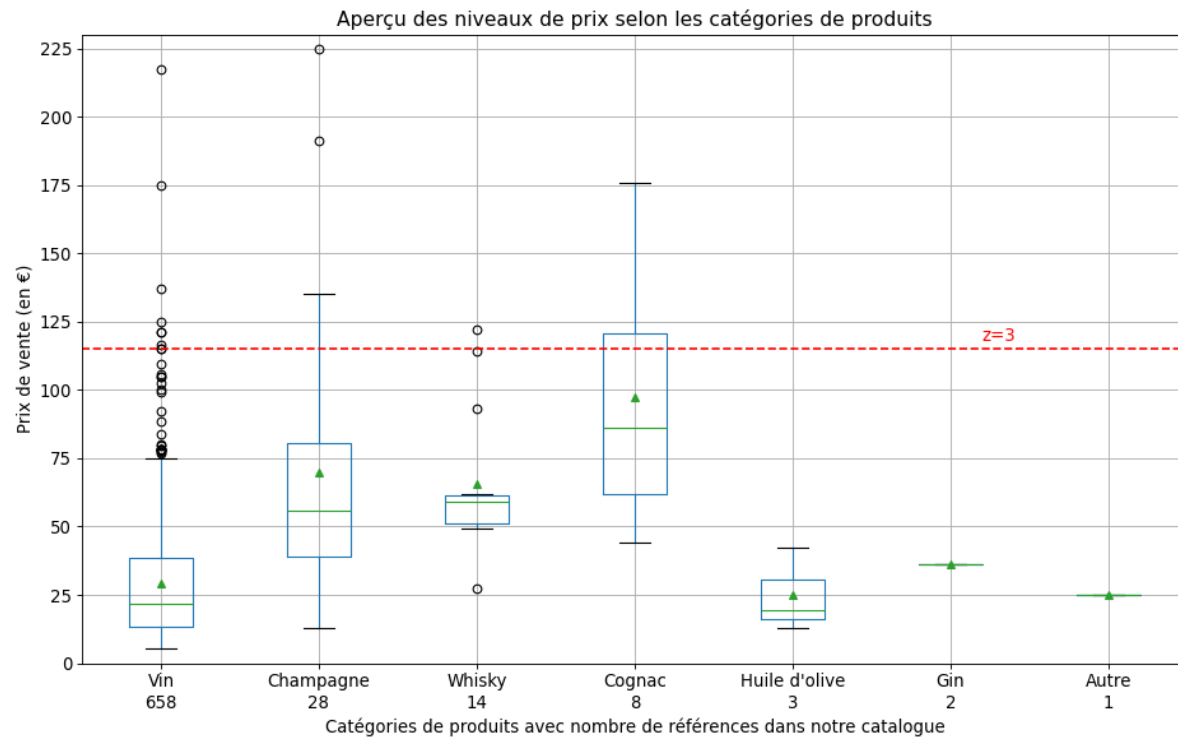
$$m = 32,33 \text{ €} \quad \text{et} \quad \sigma = 27,60 \text{ €}$$

Mesure z le nombre d'écarts-type entre le prix de l'article considéré et le prix moyen des articles du catalogue :

$$z = (\text{prix} - m) / \sigma$$

→ **13 articles du catalogue (sur 714) pour lesquels $z > 3$**

Prix selon les types de produits



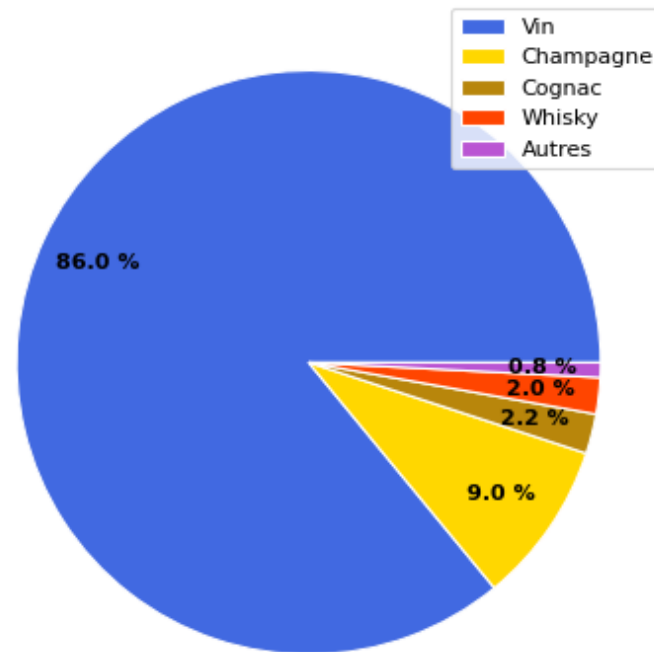
Analyse du CA par type de produits

CA total octobre : 143 680 €

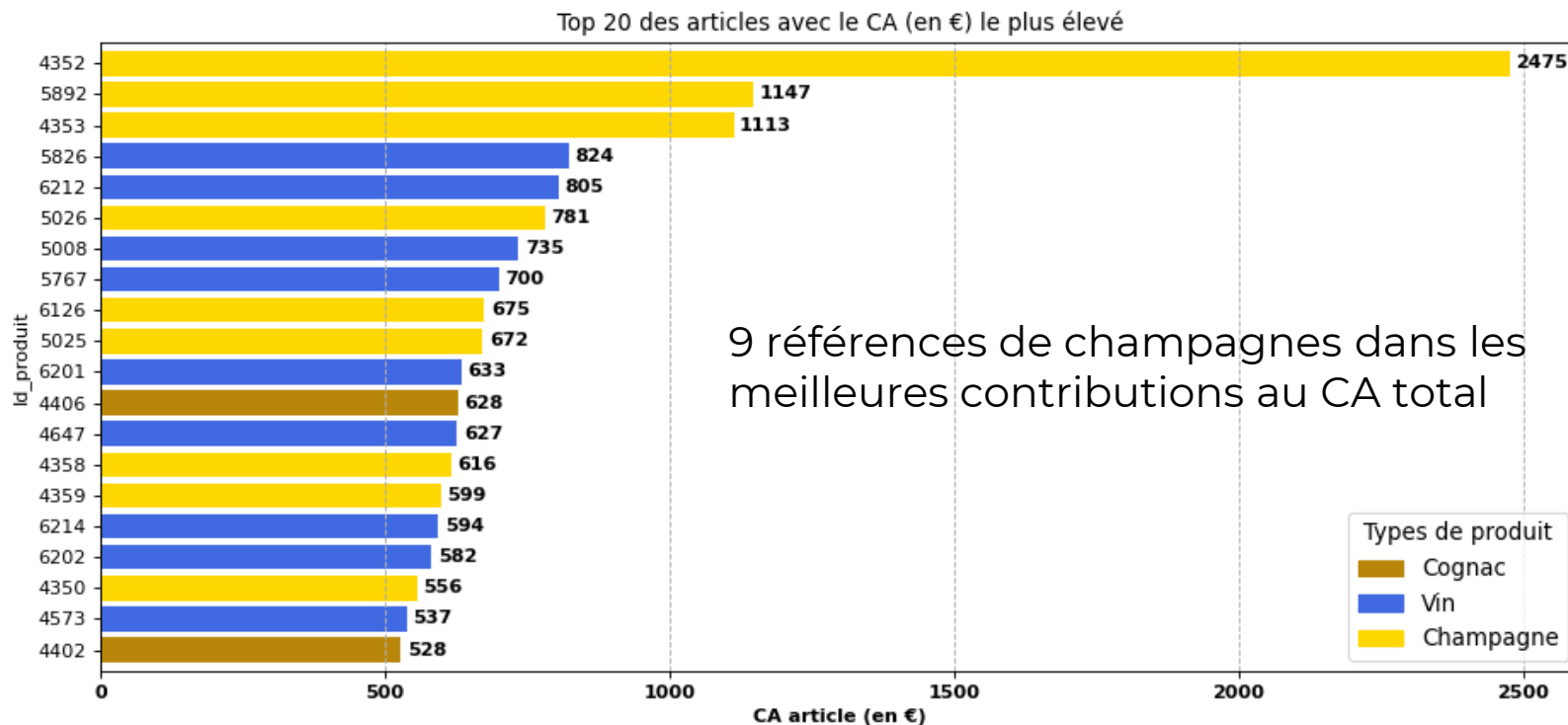
Surtout porté par les **vins** :
environ **123,5 k€**

Champagnes : environ **13 k€**

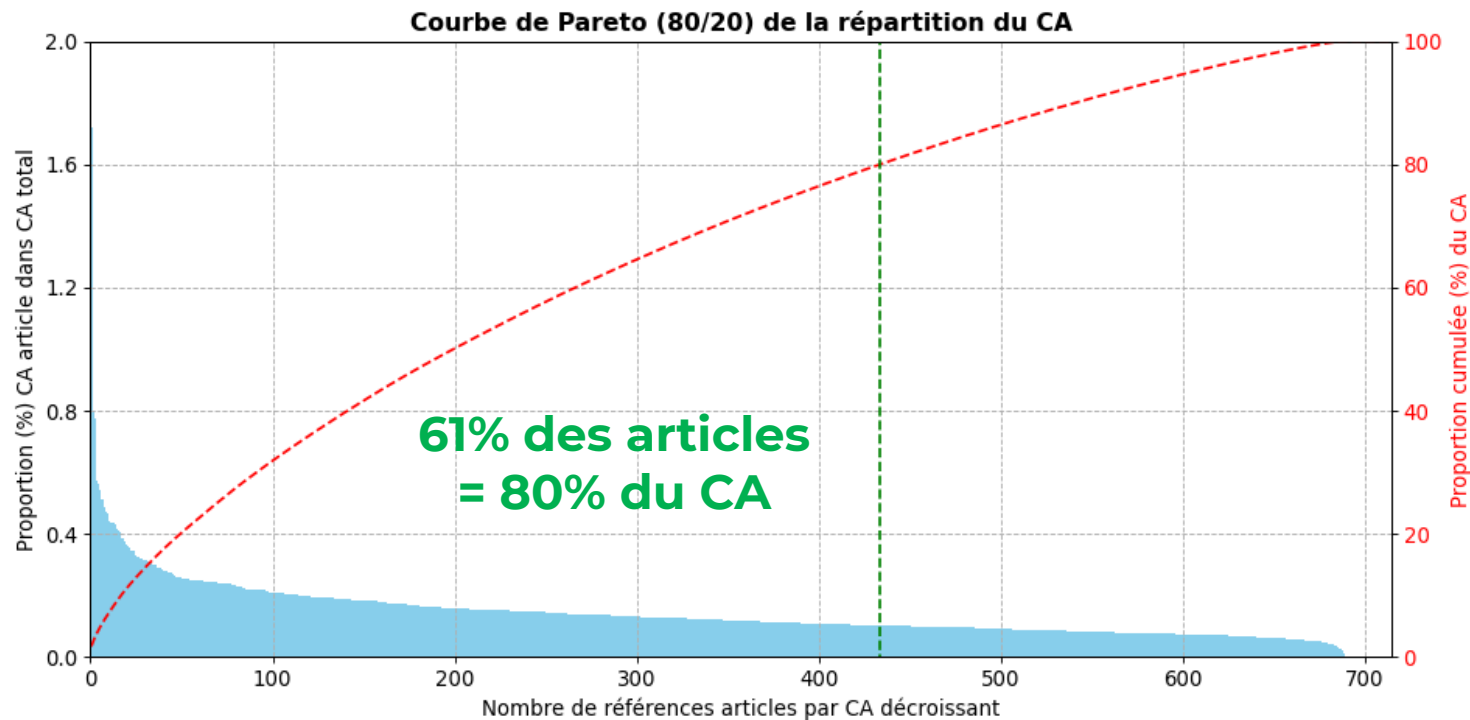
Répartition du CA (en %) par type de produits



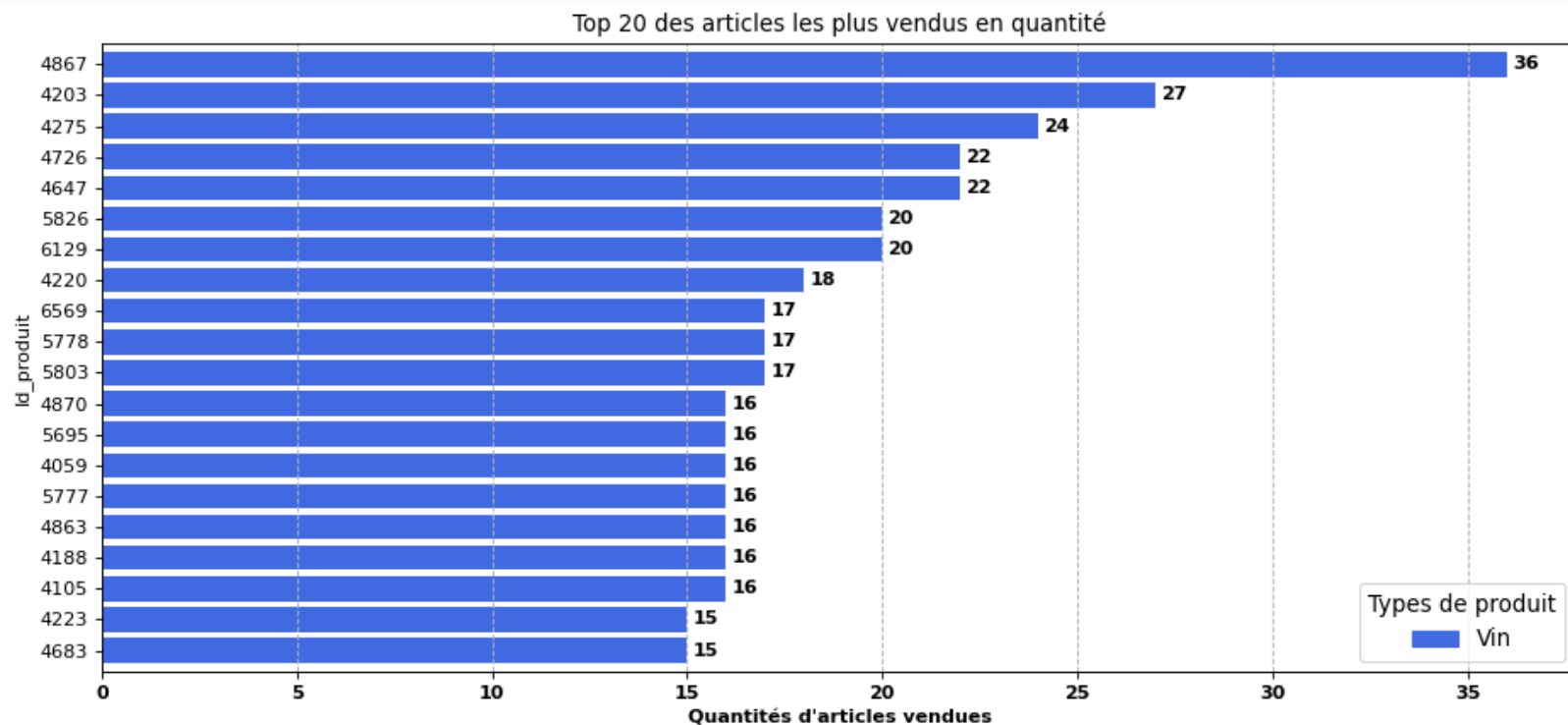
Analyse du CA : Top 20 meilleures références



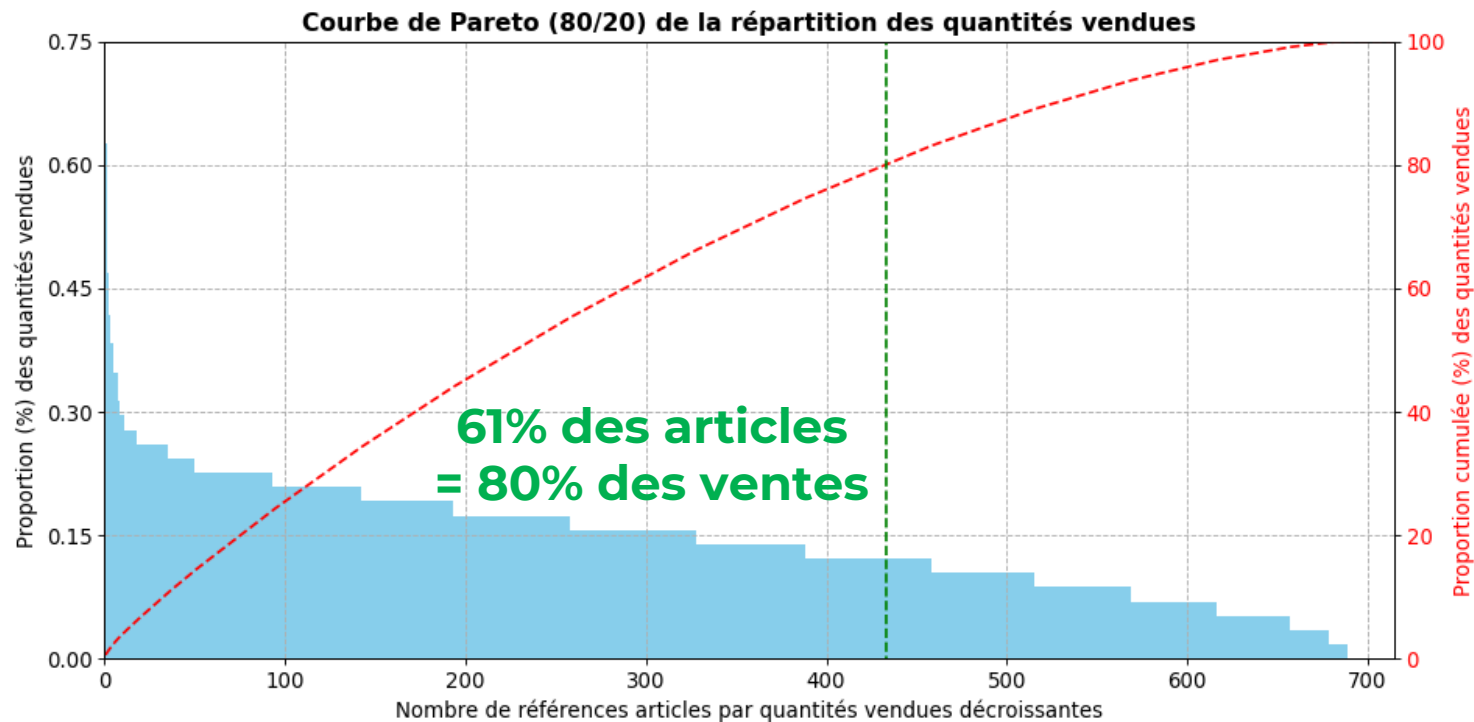
Analyse du CA : répartition 80/20



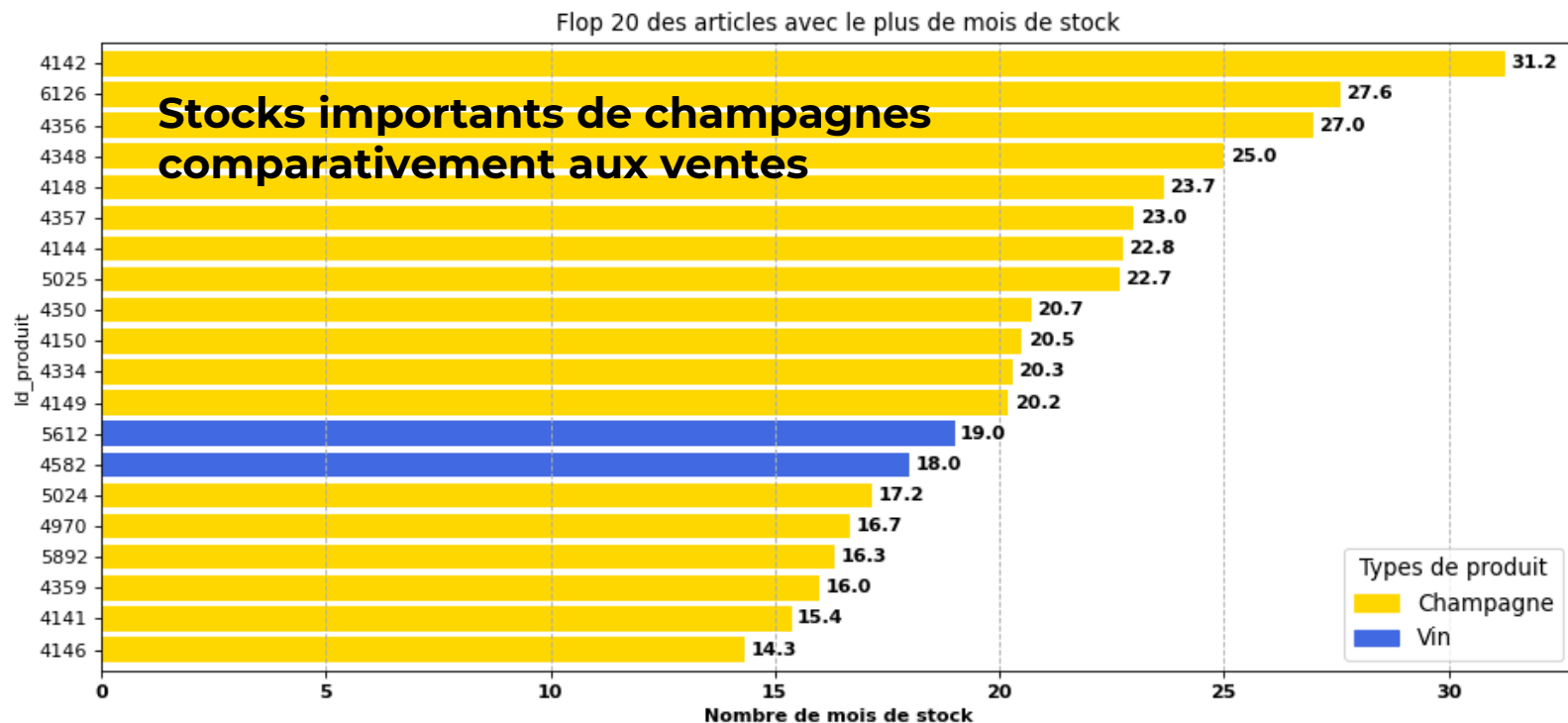
20 produits les plus vendus en quantité



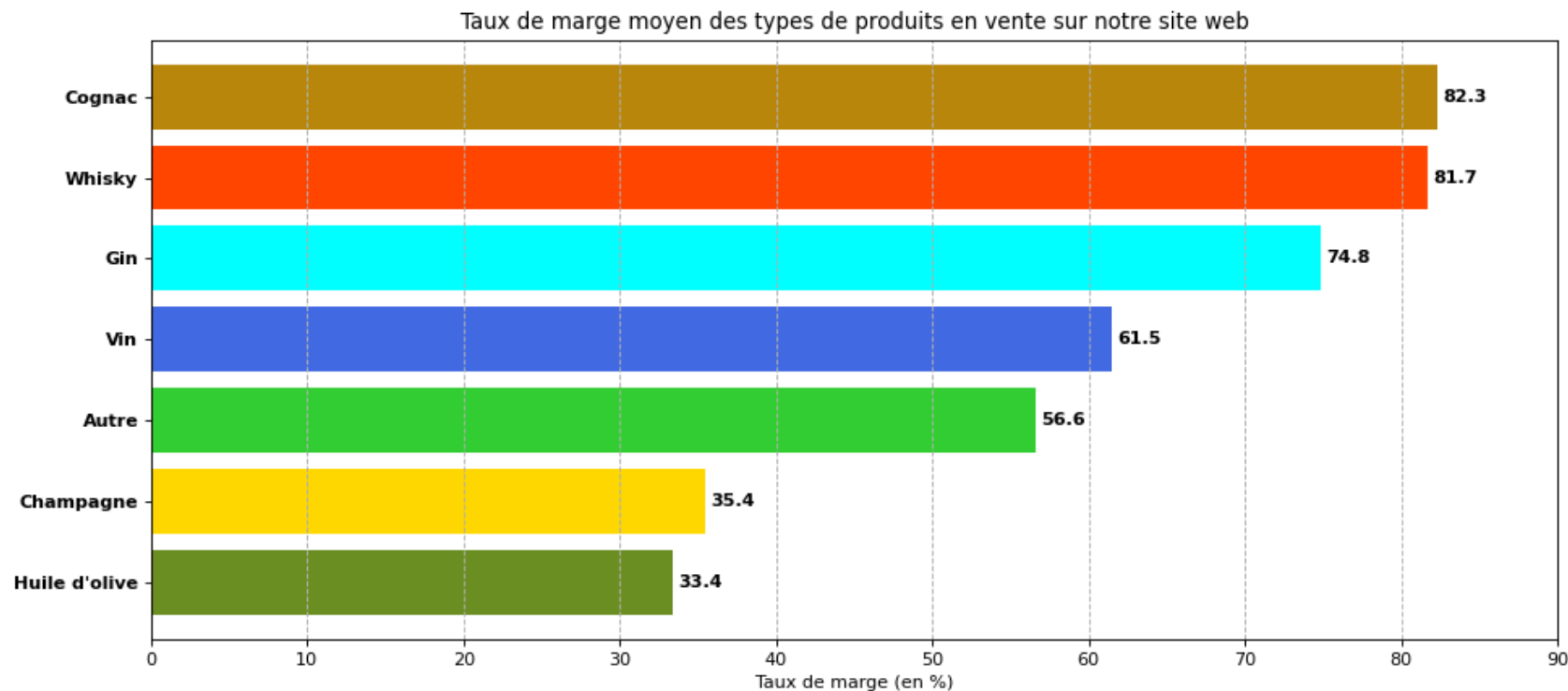
Produits les plus vendus : répartition 80/20



20 produits avec le plus de mois de stock



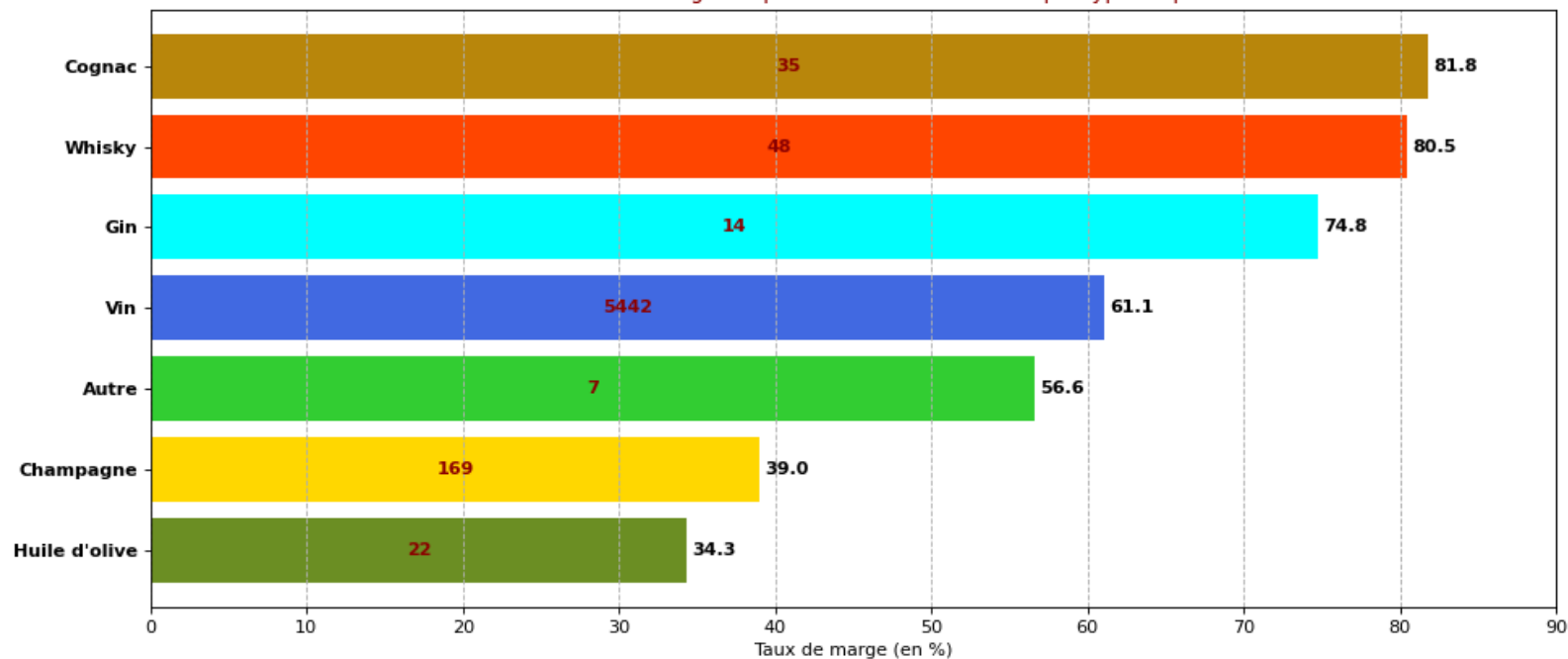
Marges moyennes - produits catalogue



Marges moyennes - produits vendus

Taux de marge moyen pondéré des types de produits vendus en octobre

à l'intérieur des rectangles : quantités vendues de chaque type de produits

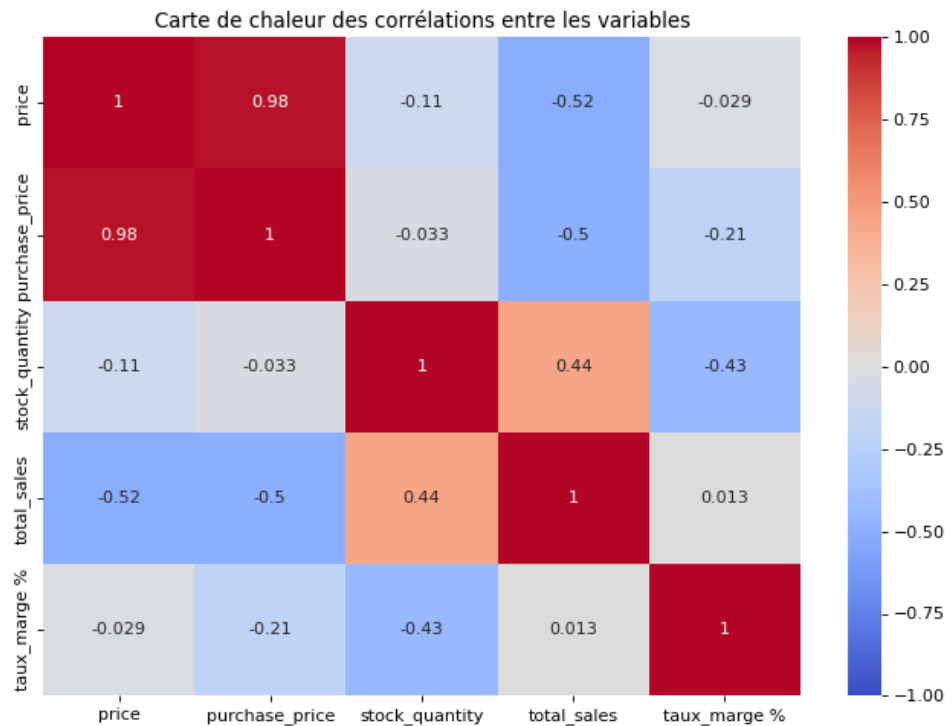


Corrélation entre variables

Forte corrélation entre prix de vente et prix d'achat

Corrélation entre stock et quantités vendues

Corrélation négative entre prix et quantités vendues

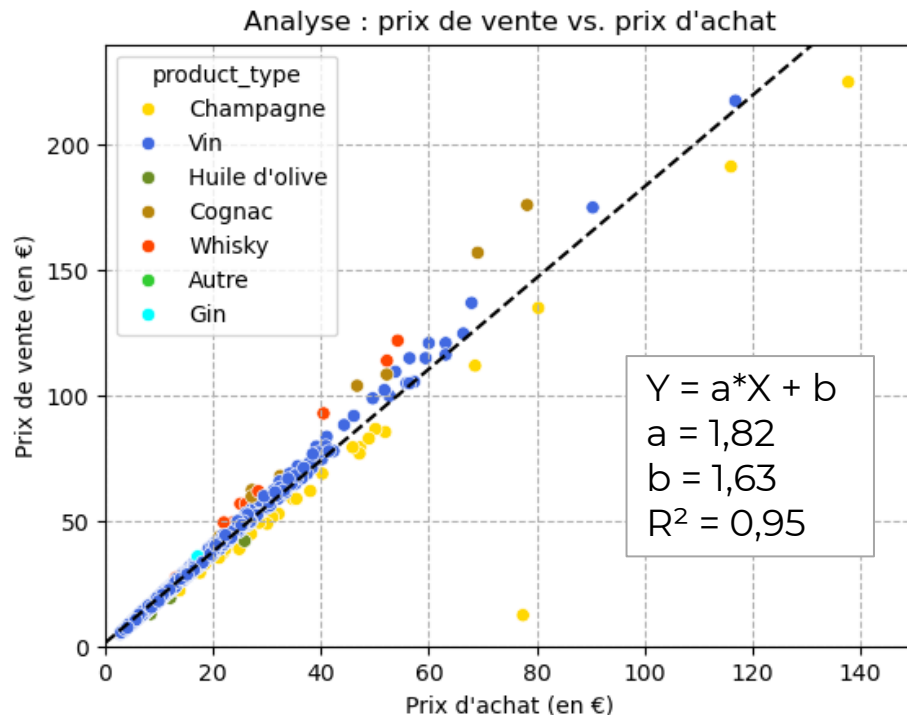


Corrélation : prix de vente vs. prix d'achat

Relation quasi-linéaire
entre prix de vente TTC
et prix d'achat ($R^2 = 0,95$)

Champagnes sous la
droite de régression

Whiskys et cognacs
au-dessus de la droite



Préconisations sur la base des analyses

La diversité des **références Vin couvre une large part du CA** et assure une bonne marge.

→ Peut-être sortir du catalogue les moins bonnes références ?

Champagnes : marge moins intéressante que celle des vins

Performances très contrastées : **être attentif aux flux de stock**

Cognacs : **très bonne marge** et quelques bonnes références

→ **Investir ce type de produits**, diversifier les références

Perspectives

Améliorer la **qualité des données du site web** :

- incohérences dans le format des *sku* ;
- *sku* manquants, quelques données anormales à corriger

Consolider ces données avec celles d'autres périodes

Bénéfice de données bien gérées :

Meilleure anticipation du marché et des stocks



Merci pour
votre attention

Avez-vous
des questions ?