



# **ANALYSE DU STOCK ET DES VENTES DU SITE BOTTLENECK**

# Analyses Exploratoires des Données

---

## Datasets:

- **ERP:** Contient des informations détaillées sur les produits, y compris les coûts et les quantités en stock.
- **Web:** Détails des produits tels que présentés sur le site web, incluant les prix, les descriptions, et les métadonnées associées.
- **Liaison:** Assure la correspondance entre les identifiants des produits dans les systèmes ERP et Web.



# Analyses Exploratoires du fichier ERP

---

## Dimensions et Composition du Dataset ERP

- **Nombre d'observations (produits) :** 825
- **Nombre de colonnes :** 6

## Traitements réalisés:

- Vérification des doublons et des valeurs null.
- Vérification de la cohérence des valeurs des colonnes 'price', 'purchase\_price', 'stock\_status' et 'stock\_quantity'.
- Correction des anomalies trouvées dans la colonne 'stock\_status' qui ne correspondait pas au 'stock\_quantity' (produits avec stock\_quantity à 0 mais avec un statut instock et d'autres avec stock\_quantity supérieure à 0 mais avec un statut outofstock)

## Remarques:

- Certains produits ont un 'purchase\_price' supérieur au 'price', ce qui n'est pas logique pour un commerce.
- Il y a des valeurs négatives dans la colonne 'price' et 'stoc\_quantity', ce qui est probablement dû à des erreurs de saisie.

# Analyses Exploratoires du fichier WEB

---

## Dimensions et Composition:

- **Nombre d'observations (produits) :** 1513
- **Nombre de colonnes :** 29

## Détail des Colonnes:

- **Colonnes avec toutes les valeurs Null :** 4 colonnes (tax\_class, post\_content, post\_password, post\_content\_filtered).

## Traitements réalisés:

- Suppression des colonnes non pertinentes pour l'analyse ou contenant uniquement des valeurs Null
- Vérification des valeurs de la colonne 'SKU':
  - Identification des 4 articles ne respectant pas la règles de codification
  - Identification des articles sans SKU
  - Suppression des lignes ne contenant que des valeurs NULL
  - Génération de 2 nouvelles valeurs SKU pour les articles indexé 1084 et 1087 en s'assurant qu'il n'y ait pas de doublons avec les SKU déjà existants
- Filtrage du dataframe sur la colonne 'post\_type' pour ne conserver que les 'product' et réduire ainsi le nombre de lignes à 716, sans aucun doublon ni valeur null dans la colonne 'SKU'

# Analyses Exploratoires du fichier LIAISON

---

## Dimensions et Composition

- **Nombre d'observations (produits) : 825**
- **Nombre de colonnes : 2**

## Détail des Colonnes:

- **id\_web (object) :** Correspond à l'identifiant de produit 'SKU' dans le fichier web. Il y a des valeurs manquantes (91 valeurs manquantes).
- **product\_id (int64) :** L'identifiant unique du produit, complet sans valeurs manquantes.

## Traitements réalisés:

- Vérification des valeurs null et des doublons
- Il y a 91 articles sans correspondances (valeurs Null dans la colonne id\_web)

# Fusion ou consolidations des données

---

- Fusion entre le **df\_ERP** et **df\_Liaison**:
  - Tous les attributs des 2 dataframe ont été conservés
  - Clé de fusion sur la colonne '**product\_id**'
  - Jointure **LEFT** utilisée car les 2 dataframes ont une colonne '**product\_id**' comportant 825 valeurs uniques non Null
- Vérifications des lignes sans correspondance et de la dimension du dataset de sortie:
  - Le tableau **df\_merge\_1** contient 825 observations et 7 colonnes.

# Fusion ou consolidations des données

---

- Fusion entre le **df\_merge\_1** et **df\_WEB**:
  - Tous les attributs des 2 dataframe ont été conservés
  - Clé de fusion sur les colonne '**id\_web**' et '**SKU**'
  - Jointure **LEFT** utilisée dans un 1<sup>er</sup> temps, puis nouvelle fusion avec une jointure **OUTER** avant de concaténer les 2 DF et d'utiliser la fonction '**drop\_duplicates**' pour afficher les 2 lignes sans correspondances avec la jointure LEFT.
- dimension du dataset de sortie:
  - Le tableau **df\_merge\_2** contient 827 observations et 19 colonnes.

# Analyses univariées du prix

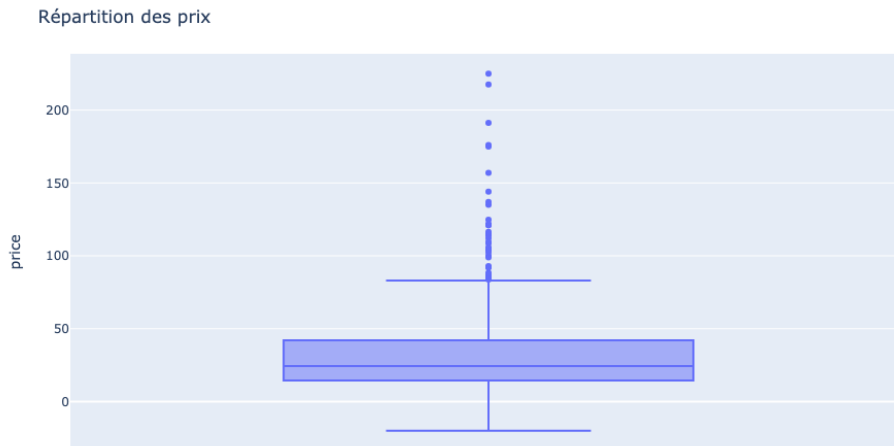
## Analyse du prix:

### La 'boite à moustaches' (Boxplot):

- **Médiane = 24,3€**
- **Quartiles**  
**Q1 = 14,5€**  
**Q3 = 42€**
- **Moustaches**  
**Upper = 83€**  
**Lower = -20€**
- **Outliers max = 225€**

### Limites Éventuelles de l'Analyse :

- L'analyse ne tient pas compte des facteurs externes qui pourraient expliquer la variation des prix, tels que la rareté, la marque, ou la qualité des produits.





# Analyses univariées du prix

**Moyenne = 32,18€**

**Ecart-type = 26,72**

**Z-Score = 1,20**

**Seuil prix pour un Z-Score de 3 est de 114€**

Seuil définit pour les outliers:

**1,5 x IQR (Q1 – Q3)**

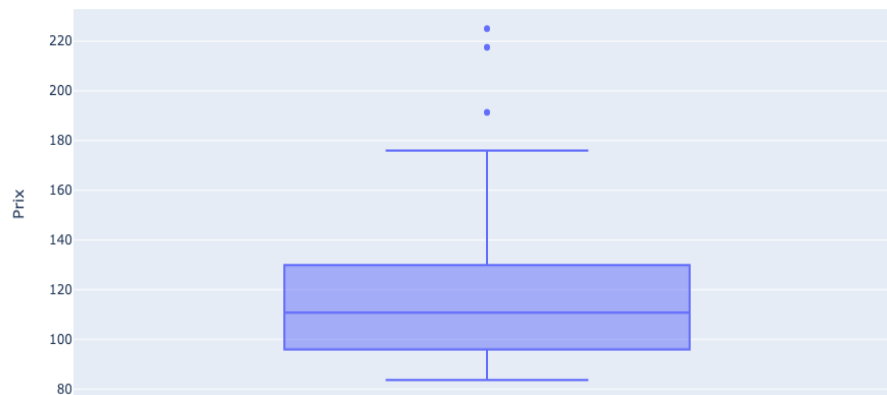
pour les **outliers en prix bas**: **-26.75€**

pour les **outliers en prix haut** : **83.25€**

Le nombre d'articles considérés comme 'Outliers' est donc de **36, soit 4% de l'ensemble du catalogue.**

Ces articles sont des **vin et spiritueux haut de gamme dont la qualité/rareté peut justifier le prix.**

Boxplot des prix des articles outliers



Médiane

**110,8€**

Quartiles

**Q1 = 96€**

**Q3 = 129,9€**

Moustaches

**Lower = 83,7€. Upper = 176€**

Outliers

**max = 225€**

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

## Analyse des ventes en CA et quantité:

**CA du mois d'octobre = 143 680,1€**

Nombre d'articles nécessaires pour atteindre  
**80% du CA :**

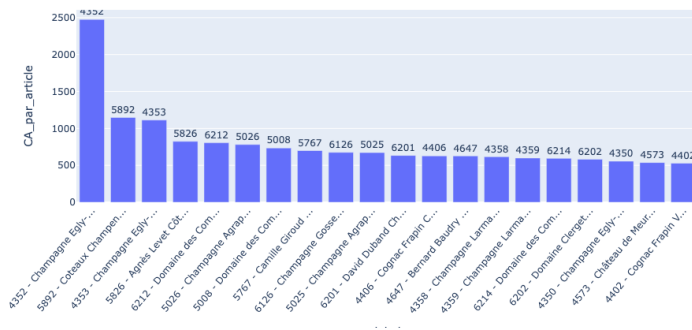
**434 soit 52,48% du catalogue**

**Total de ventes d'octobre = 5 678**

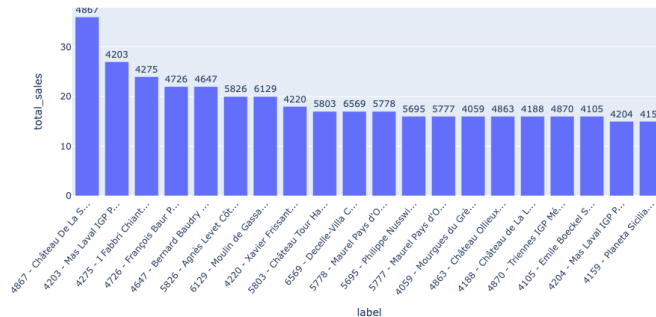
Nombre d'articles nécessaires pour atteindre  
**80% du total des ventes :**

**424 soit 51,27% du catalogue.**

Top 20 des articles en chiffre d'affaires



Top 20 des articles par nombre de ventes

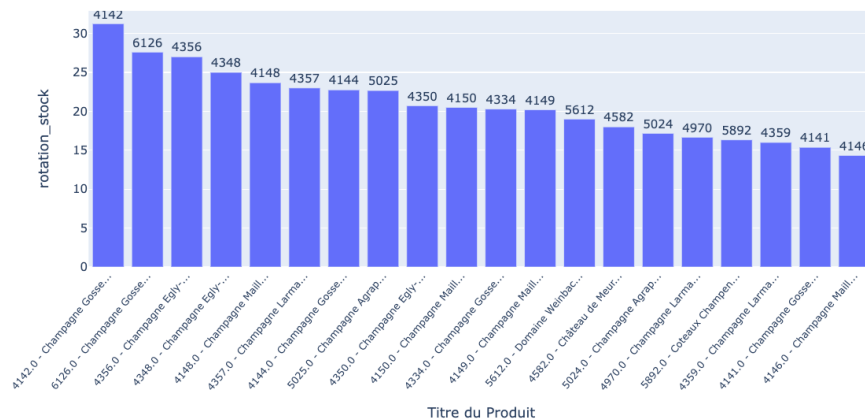


## Analyse des stocks:

La valorisation du stock actuel est de **531 628,80€** pour un total de **17 811** articles.

Le nombre d'articles actuellement 'out of stock' est de seulement 92, soit 0,52% de notre catalogue.

Flop 20 des produits avec le plus de stock



## Analyse du taux de marge

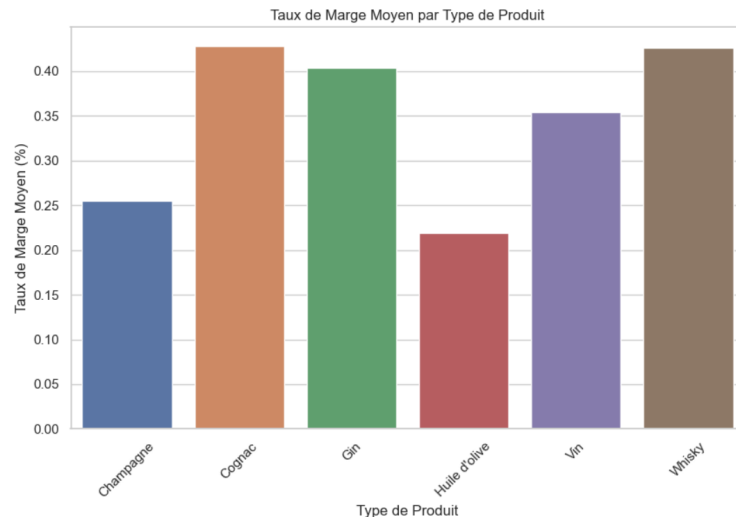
**La marge moyenne est de 35,70%**

**Le taux de marge mini est de 19,60%**

**Le taux de marge maxi est de 167,80%**

*\*A noter que 4 articles ont un prix de vente HT < au prix d'achat donc une marge négative.*

*Ces 4 articles n'ont pas été pris en compte dans le calcul des marges.*



product_type	tx_marge_moyen
Champagne	0.255089
Cognac	0.427788
Gin	0.404167
Huile d'olive	0.218731
Vin	0.354294
Whisky	0.426238

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

### Analyse de la rentabilité par produit

- **Champagne** et **Huile d'olive** présentent des taux de rotation très faibles, ce qui indique un potentiel excès de stock par rapport à la demande.
- **Gin** et **Whisky** montrent des taux de rotation élevés, suggérant une bonne adéquation entre le stock et la demande.
- **Vin** a la plus grande marge nette totale et un taux de rotation raisonnable, ce qui en fait un produit clé pour le chiffre d'affaires.

	Marge Nette Totale	Ventes Totales	Stock Total	Taux de Rotation
product_type				
<b>Champagne</b>	2593.08	169.0	2894.0	0.058397
<b>Cognac</b>	1083.19	35.0	99.0	0.353535
<b>Gin</b>	162.96	14.0	11.0	1.272727
<b>Huile d'olive</b>	89.20	22.0	158.0	0.139241
<b>Vin</b>	34919.16	5383.0	13469.0	0.399658
<b>Whisky</b>	976.82	48.0	85.0	0.564706

### Prix de vente theorique

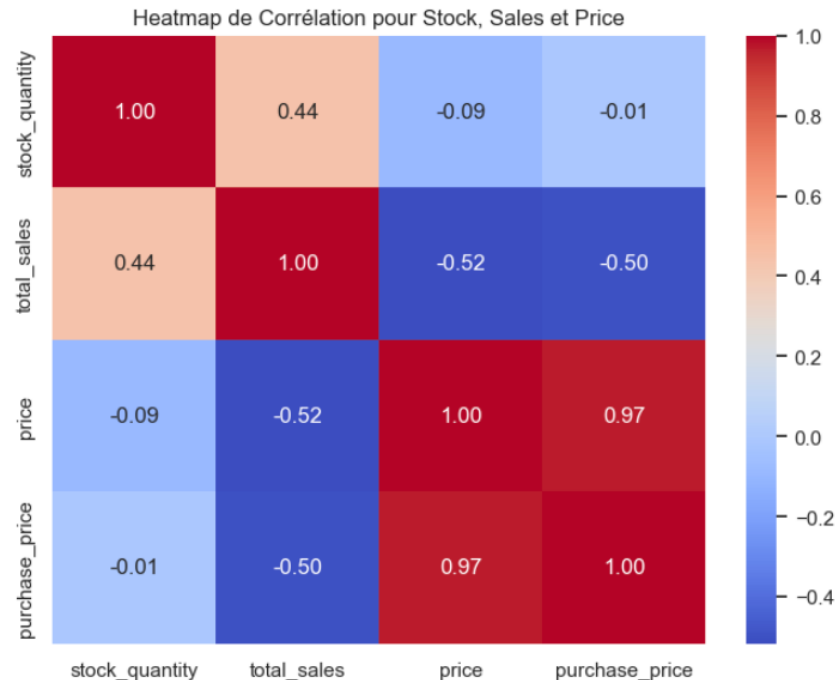
product_type	purchase_price	tx_marge_moyen	prix_vente_theorique_moyen
Champagne	42.930370	0.255089	53.881433
Cognac	44.286250	0.427788	63.231360
Gin	17.160000	0.404167	24.095500
Huile d'olive	15.450000	0.218731	18.829397
Vin	15.145106	0.354294	20.510933
Whisky	29.937143	0.426238	42.697496

# Analyses complémentaires

## CA, quantités, stocks, taux de marge et correlations

### Corrélations

- **Stock Quantity et Total Sales:** Corrélacion positive moyenne (0.44). Cela suggère que lorsque la quantité de stock augmente, les ventes totales ont tendance à augmenter également, mais pas aussi fortement que ce serait le cas avec une corrélation parfaite.
- **Total Sales et Price:** Corrélacion négative assez forte (-0.52). Cela suggère que lorsque le prix augmente, les ventes totales ont tendance à diminuer, ce qui est intuitif car les produits plus chers pourraient être moins achetés.
- **Total Sales et Purchase Price:** Corrélacion négative assez forte (-0.50). Ceci est similaire à la corrélation entre les ventes totales et le prix, indiquant que les ventes peuvent baisser lorsque le prix d'achat augmente.
- **Price et Purchase Price:** Très forte corrélation positive (0.97). Cela suggère que le prix de vente et le prix d'achat sont presque toujours proportionnels, ce qui signifie que lorsque le prix d'achat augmente, le prix de vente fait de même.



# Actions pour la suite

---

## 1. Uniformisation des Références Produits :

Établir une clé de référence unique commune entre tous les systèmes (ERP, site Web, table de liaison) pour faciliter les mises à jour et l'intégration des données.

## 2. Automatisation des Mises à Jour :

Mettre en place un processus d'intégration automatique des données entre l'ERP et le site Web pour éviter les mises à jour manuelles, ce qui réduit le risque d'erreurs et économise du temps.

## 3. Contrôles de Cohérence :

Implémenter des contrôles de cohérence et de validation pour s'assurer que les informations sur les stocks et les prix sont à jour et exactes entre les différentes plateformes.

## 4. Gestion des Stocks en Temps Réel :

Assurer une synchronisation en temps réel ou quasi-réel des niveaux de stock pour éviter les ruptures de stock ou les ventes en surstock sur le site Web.



# Point sur les compétences apprises

---

- *Qu'est-ce qui s'est bien passé pour vous dans ce travail de nettoyage ?*

L'analyse univarié/multivarié est très intéressante à mettre en pratique.

- *Qu'est-ce que vous avez trouvé le plus difficile ?*

Quelques difficultés rencontrés lors du merge entre les DF avec la gestion des valeurs Null

- *Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ?*

Nettoyage des données