

A high-contrast photograph of a glass of red wine being poured. The glass is on the left, tilted, with a stream of red wine falling into it. The background is black, with a large, stylized red splash or stain on the right side. The text is overlaid on the right side of the image.

# Rapprochement des données et analyse des ventes chez BottleNeck

Théo ZWAHLEN  
DATA ANALYST

23/10/2024



# Datasets & Caractéristiques

## Sources de données :

1. **ERP** : Données sur les produits, prix, stock, et prix d'achat.
2. **Site Web** : Informations sur les ventes, titres des produits, et dates de publication.
3. **Table de liaison** : Permet de connecter les références produits entre l'ERP et le site Web via les identifiants SKU et Product ID.

## Caractéristique des données :

- Qualité des données : valeurs aberrantes, prix et stock négatif / doublons et SKU manquants ou non conformes
- Incohérences détectées : Problèmes de correspondance entre les SKU du site Web et les product\_id de l'ERP

## Traitements réalisés :

- Nettoyage des données : Correction des SKU manquants ou non conformes, doublons, prix et stock négatifs
- Suppression des colonnes inutiles
- Features Engineering : Création de colonnes calculées comme ca\_par\_article et rotation\_stock



# Fusion ou consolidations des données

## Choix des attributs :

- Nous avons sélectionné les colonnes essentielles pour l'analyse, telles que product\_id , sku , price , totale\_sales et stock\_quantity
- Les informations cruciales comme les prix d'achat et la rotation\_stock ont également été incluses pour des analyses supplémentaires.

## Clés utilisées :

- **product\_id** a été utilisé comme clé principale pour la jointure entre le fichier **ERP** et la **table de liaison**.
- **sku** (identifiant unique des produits sur le site Web) a permis de faire le lien entre la **table de liaison** et les données du **site Web**.

# Fusion ou consolidations des données

## Vigilances particulières :

- **L'alignement des données** : s'assurer que chaque produit est bien relié entre l'ERP et le site Web.
- **Identification des lignes sans correspondance** : Utilisation de la jointure externe (how='outer') pour détecter les produits manquants.
- **Conséquences pour l'analyse** : L'absence de correspondance entre les fichiers peut fausser les résultats d'analyses. Par conséquent, il est crucial d'identifier et de comprendre l'impact de ces lignes non correspondantes.

## Difficultés rencontrées :

- **Gestion des SKU manquants et incohérences** : 111 lignes sans correspondance ont été identifiées entre les fichiers ERP et Web.
- **Interprétation des données left\_only** : La colonne \_merge a révélé des produits présents uniquement dans df\_merge mais pas dans df\_web.
- **Impact sur les décisions** : Les lignes sans correspondance entre les sources de données nécessitent une analyse plus approfondie pour éviter des décisions basées sur des données incomplètes.

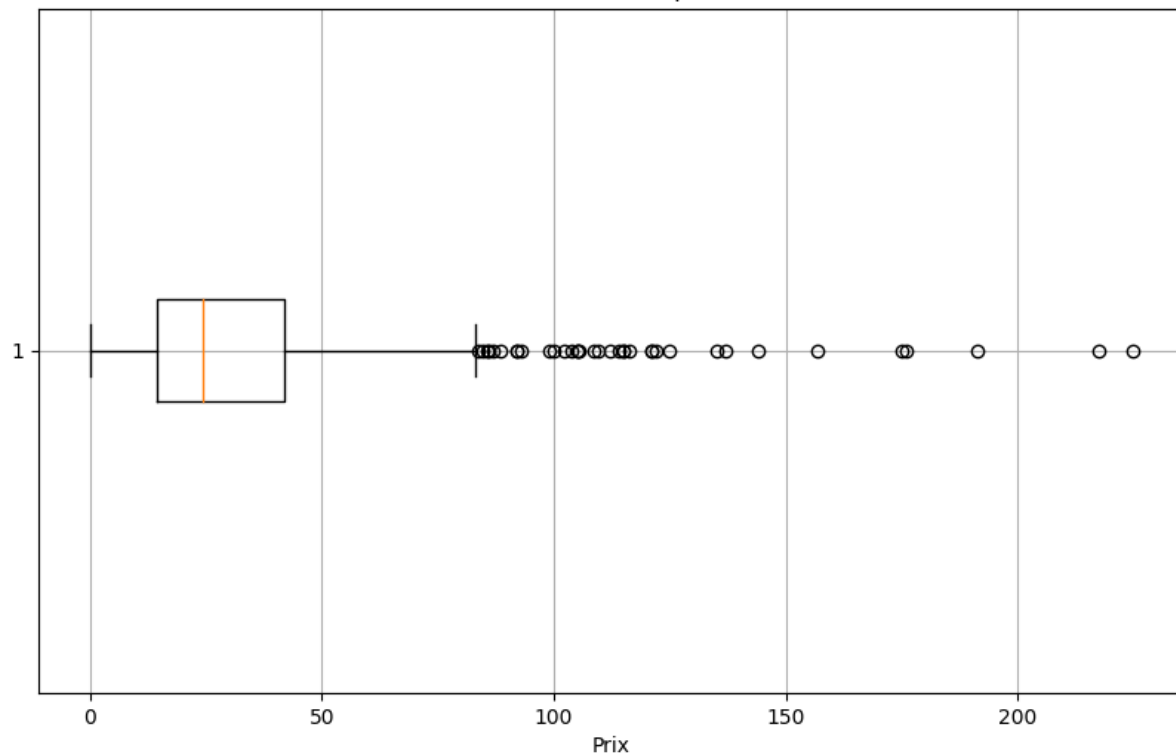
# Analyses univariées du prix

## Méthodes statistiques employées :

- **La moyenne** des prix a été calculée à **32,23 euros**,
- **L'écart-type** est de **26,65 euros**.
- une grande variabilité dans les prix,
- **Z-score** : Calcul du Z-score pour identifier les valeurs extrêmes en fonction de la moyenne.
- Seuil de 3 utilisé pour isoler les prix significativement élevés.
- Résultat : un seuil de 114 € pour les valeurs extrêmes
- **Méthode de l'intervalle interquartile (IQR)** :
  - Calcul des bornes de  $Q1 - 1.5 * QR$  et  $Q3 + 1.5 * QR$  pour détecter les valeurs aberrantes.
  - Borne supérieure fixée à 83,25 €.
- **Résultats des valeurs aberrantes** : Les produits dont le prix dépasse cette borne supérieure (83,25 euros) ont été identifiés comme valeurs aberrantes.
- Un total de **36 produits** ont été classés comme des outliers par cette méthode.

# Analyses univariées du prix

Boîte à Moustache de la Répartition des Prix

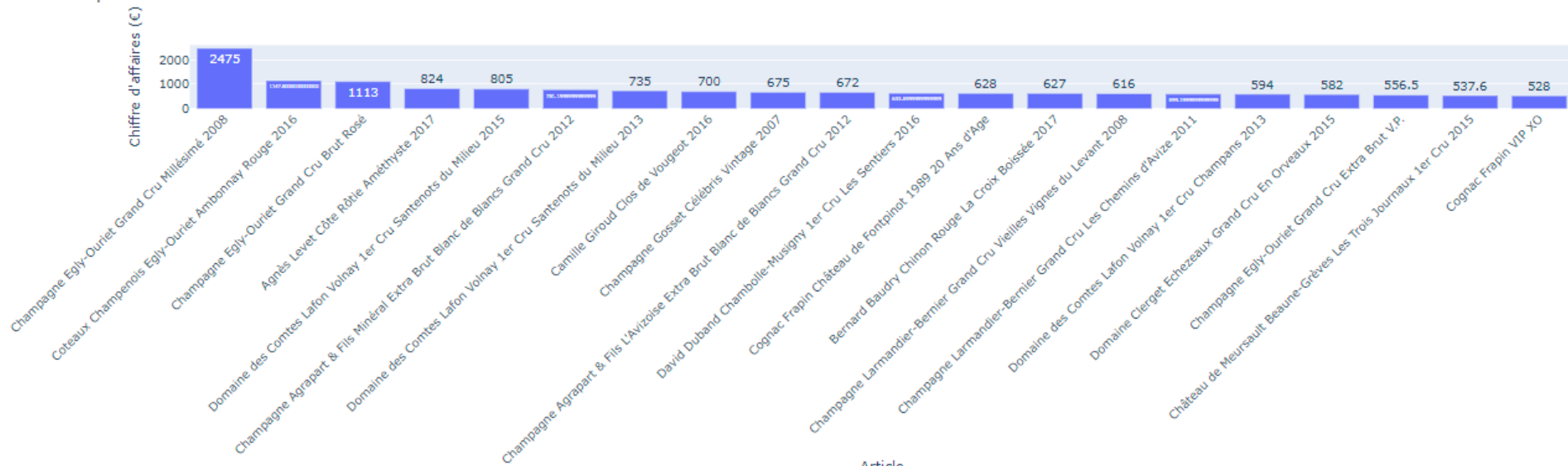


# Analyses complémentaire :

## Analyse du Chiffre d'Affaires par Produit

- Calcul du **chiffre d'affaires (CA)** par produit et total.
- Chiffre d'affaires total : 143 680,10 €.

Top 20 des articles en fonction du chiffre d'affaires





# Analyses complémentaire :

## Répartition 20/80 du CA

### **Explication de la règle du 20/80 :**

- Le principe du 20/80, également connu sous le nom de Loi de Pareto, stipule que 20 % des causes sont souvent responsables de 80 % des effets. En commerce, cela signifie que généralement, 20 % des produits génèrent environ 80 % du chiffre d'affaires.

### **•Application dans notre analyse :**

- Dans notre cas, nous avons constaté que **52,61 %** de nos articles génèrent **80 %** du chiffre d'affaires total.

### **•Implication pour l'entreprise :**

- Cela signifie que plus de la moitié de notre catalogue contribue majoritairement aux revenus, ce qui représente une dispersion des ventes un peu plus élevée que le modèle typique du 20/80.

- Identifier et concentrer nos efforts marketing et de gestion sur ces produits pourrait améliorer la performance et l'efficacité des ressources.





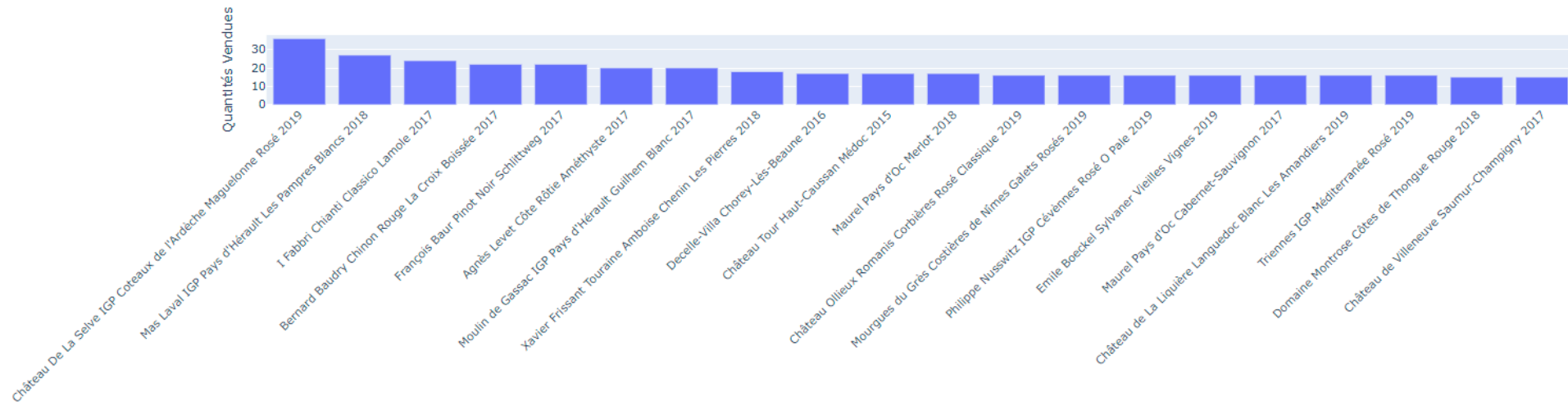


# Analyses complémentaire :

## Analyse des Ventes en Quantités

- Ces informations peuvent orienter les décisions stratégiques sur les stocks et les opérations marketing pour les produits populaires.

Top 20 des Articles en Quantités Vendues

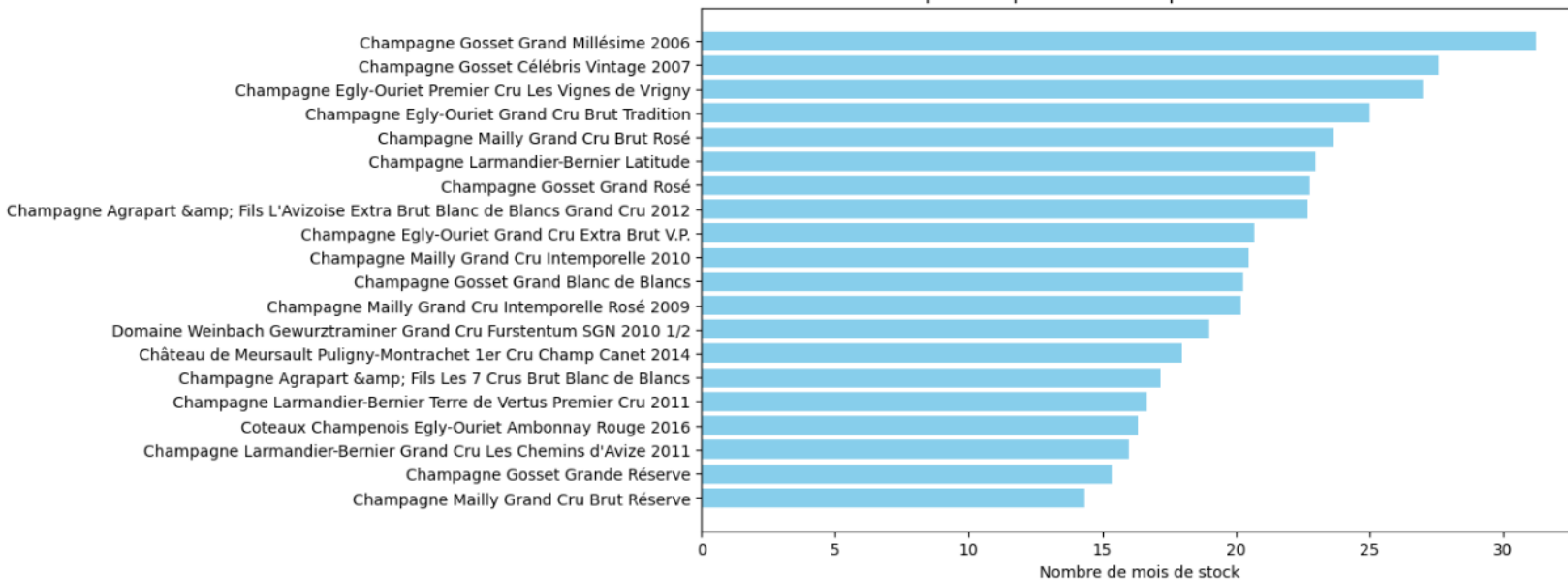


# Analyses complémentaire :

## Analyse des Stocks et Rotation de Stock

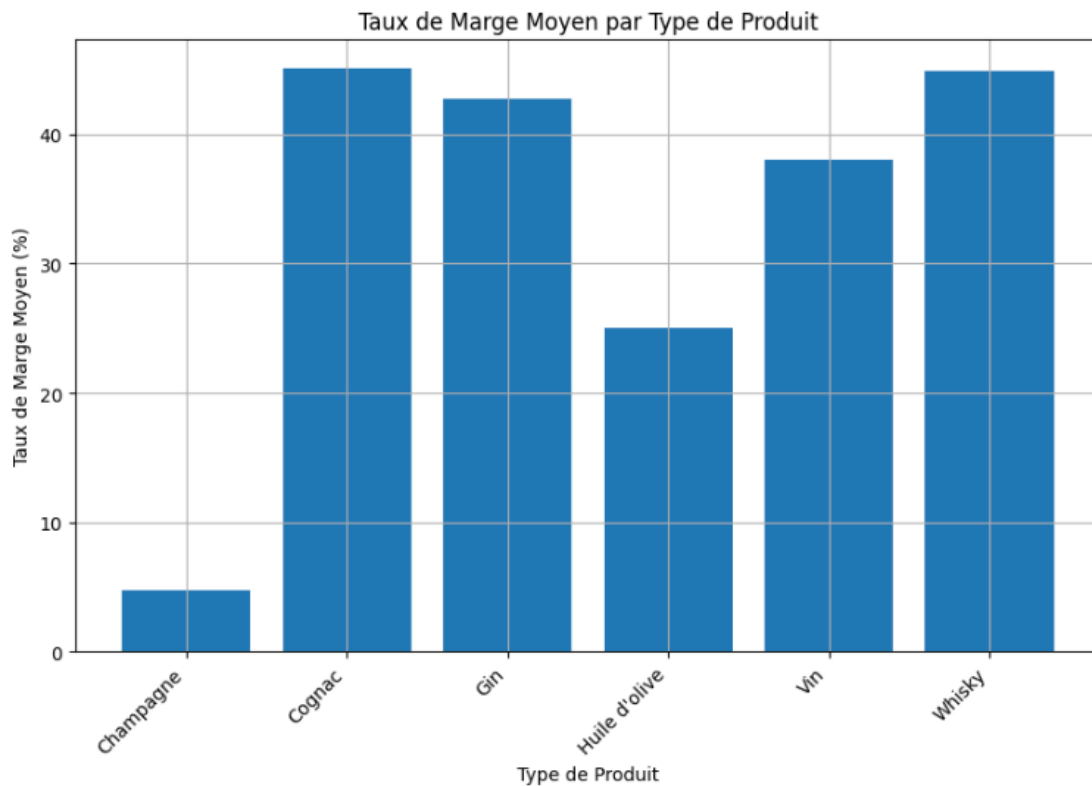
- **Produits en stock** : 17822€
- **Valorisation totale des stocks** : 531 946,20 €

Top 20 des produits avec le plus de mois de stock



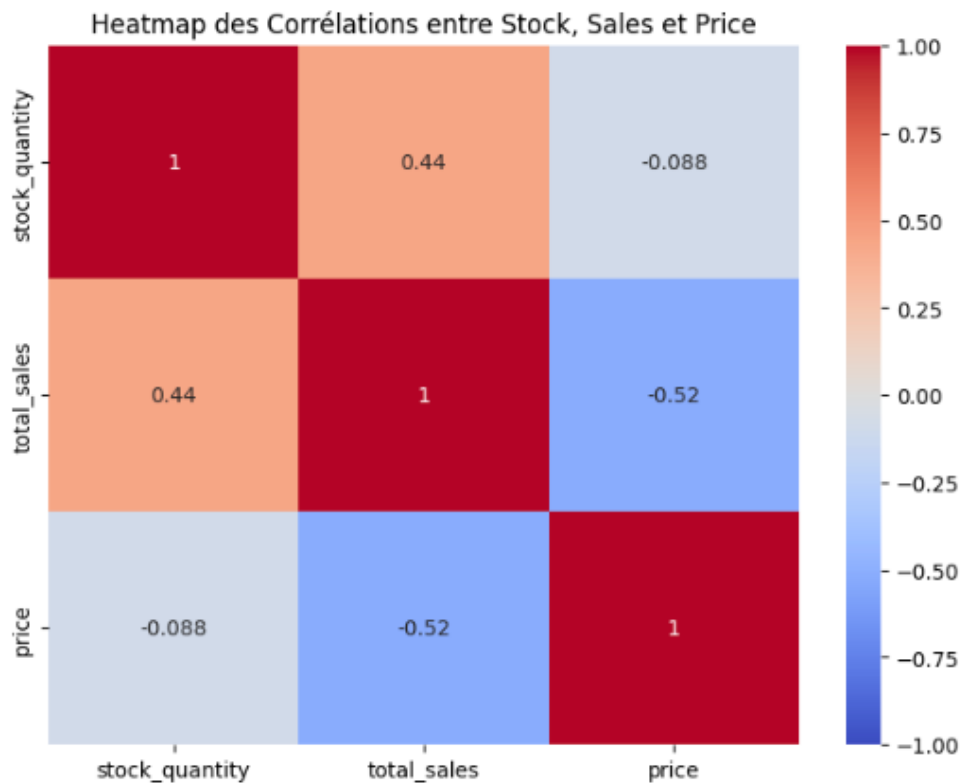
# Analyses complémentaire :

## Taux de Marge par Type de Produit



# Analyses complémentaire :

## Corrélations entre les Variables Stock, Ventes et Prix





# Analyses complémentaire :

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

### Limites éventuelles de l'analyse :

- Les valeurs aberrantes dans le prix et les quantités peuvent influencer les résultats de la corrélation.
- Les analyses sont basées sur une période spécifique, ce qui peut limiter la généralisation des conclusions sur d'autres périodes ou saisons.



# Actions pour la suite

## Amélioration des données :

- Mettre en place des contrôles de qualité pour détecter les incohérences et les valeurs aberrantes, comme les prix ou stocks négatifs, dès l'intégration des données.
- Standardiser le format des SKU pour éviter les erreurs de saisie.

## Optimisation des ventes et du stock :

- Analyser les articles à faible rotation et proposer des actions marketing pour stimuler les ventes (ex. promotions ciblées).
- Réduire les stocks des produits à faible rotation afin d'optimiser la gestion des inventaires.

## Suivi des indicateurs clés :

- Mettre en place un tableau de bord pour suivre en temps réel les indicateurs clés (CA, quantités vendues, taux de marge).
- Intégrer des alertes pour notifier lorsque certains seuils sont atteints (ex. seuils critiques pour le stock).

# Point sur les compétences apprises

## **Points positifs :**

Bonne gestion du nettoyage des données, correction des erreurs et ajustements sur les prix et stocks négatifs.

## **Défis rencontrés :**

Difficultés dans la correspondance entre SKU du site Web et Product ID de l'ERP.  
La résolution des incohérences dues aux valeurs manquantes et aux doublons.

## **Axe d'amélioration :**

Approfondir les compétences en analyse multivariée et en visualisation des corrélations.  
Développer une maîtrise plus avancée des outils statistiques.