



Optimisation de la gestion des données de Bottleneck avec Python

SERE Julien

Data Analyst Bottleneck

08/12/2024

Sommaire

- Analyses exploratoires des données
- Fusion ou consolidation des données
- Analyses
- Actions pour la suite

Analyses Exploratoires des Données

- 3 fichiers de données

Web : Informations web (1513 observations)

ERP : Informations des produits dans l'ERP (825 observations)

Liaison : Avec références produits des 2 tables (825 observations)

- Traitements réalisés
 - Corrections d'incohérences sur les statuts du stock
 - Corrections de prix négatifs
 - Erreurs de stocks (valeurs négatives)
 - Suppressions de colonnes inutiles à l'analyse
 - Suppressions de lignes aux valeurs nulles
 - Suppressions de doublons lié à la présence d'une photo pour chaque article => les produits étaient doublés
- Choix de traitement au cas par cas (règles de codification)

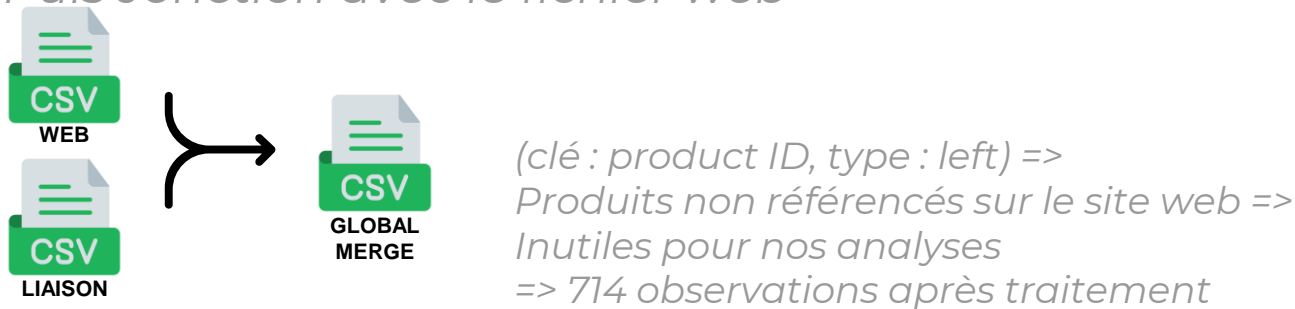


Fusion ou consolidations des données

- *Jonction du fichier ERP & Liaison*



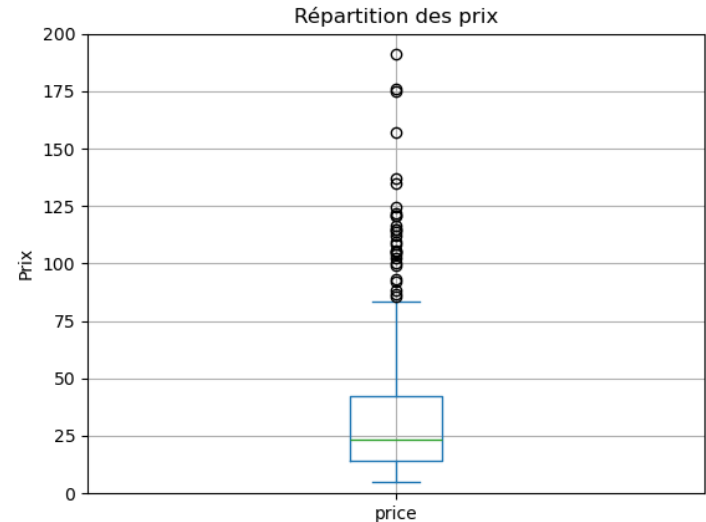
- *Puis Jonction avec le fichier web*



Analyses univariées du prix

Boîte à moustache

- Médiane basse : 32 €
- Ecart type : 27.1€
- Valeurs atypiques nombreuses => répartition asymétrique avec un étirement vers les prix élevés.
- Identification des outliers :
- La méthode de l'intervalle interquartile est plus adaptée car les données suivent une distribution asymétrique comme on a pu l'observer (pas de loi normale)



Analyses

CA, quantités, stocks, taux de marge et correlations

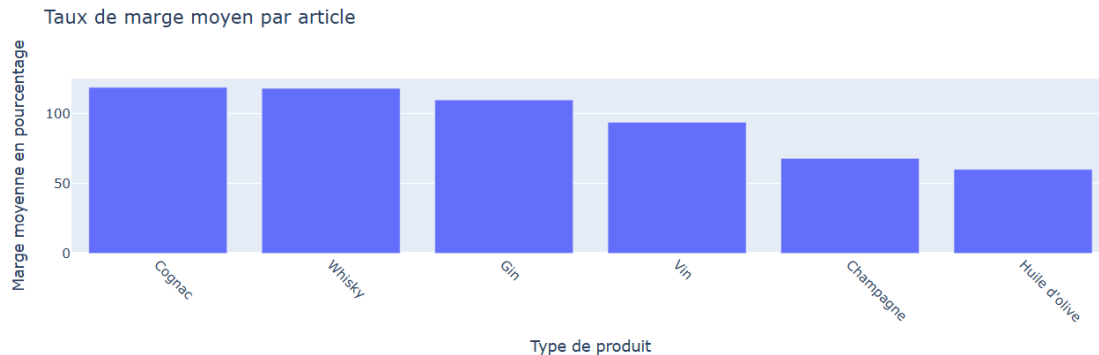
● **Mesures :**

C.A par article (Proportion des articles représentant 80% du CA s'élève à 60.78 %

C.A en quantité (Proportion des articles représentant 80% du CA en quantité s'élève à 2.80 %)

Marge minimum : 55 %

Marge maximum : 129%



Analyses

Corrélations principales

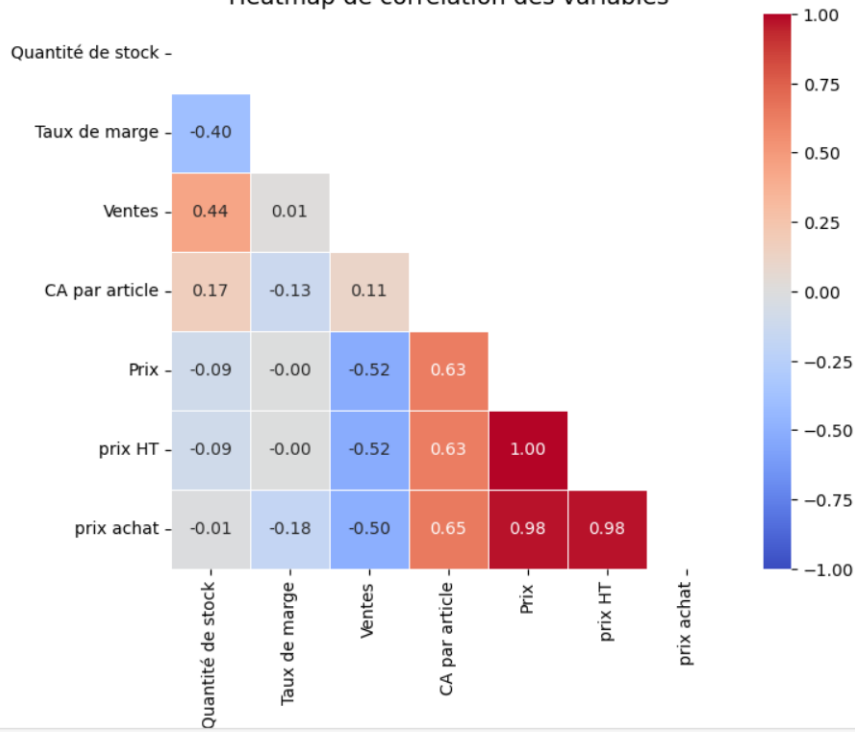
● **Mesures :**

Augmentation du prix => baisse modérée des ventes en quantité

Augmentation des stocks => Plus de ventes et C.A sur certains produits

Cependant, les produits les plus chers pourraient être légèrement moins stockés.

Heatmap de corrélation des variables

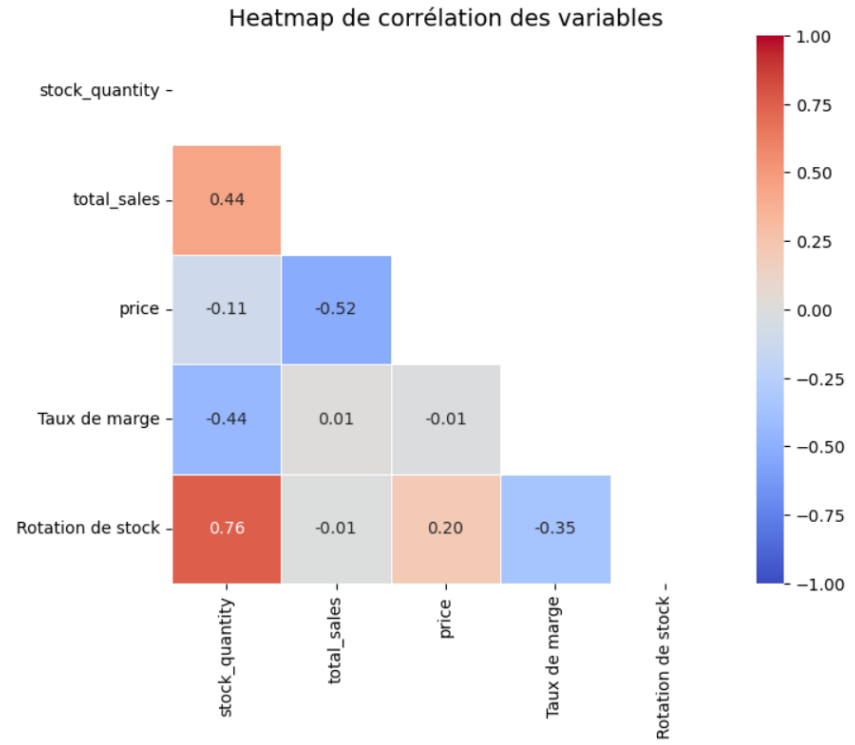


Analyses Corrélations supplémentaires



- **Mesures supplémentaires :**

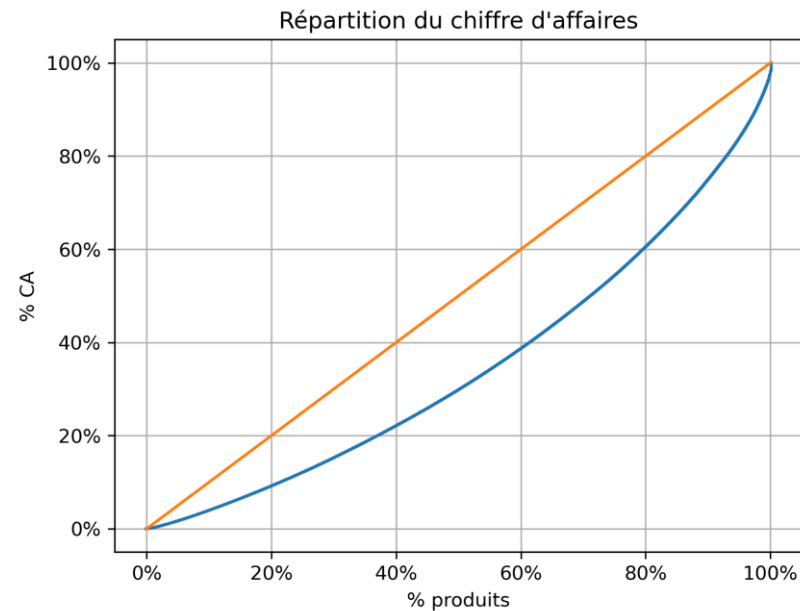
Un taux de marge élevé est associé à une rotation de stock plus lente.



Pour aller plus loin... Répartition du C.A



- *Courbe de Lorenz*
50 % du chiffre d'affaires est réalisé par 35 % des produits.
- *Indice de Gini : 0.30*
Répartition du chiffre d'affaires est inégalitaire.



Actions pour la suite

Traitement des données :

- *Unification les articles dans les fichiers internet afin de ne garder qu'une ligne produit par article*
- *Amélioration des contrôles internes sur la saisie et le suivi des stocks*
- *Utiliser des algorithmes tels que Isolation Forest pour détecter des outliers multivariés*



Stratégie :

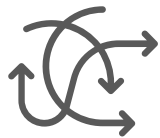
- *Créer un système de requête par internet pour contrôler et valider les outliers via API ou WebScraping*
- *Segmentation des analyses des données par type de produit*
- *Amélioration de la gestion des stocks : la disponibilité des produits est un facteur important dans les ventes.*
- *Etudes des stratégies de tarification (modèles dynamiques...)*



Point sur les compétences apprises



Analyse exploratoire des fichiers
Identification des valeurs atypiques ou incohérentes



Export des graphiques et utilisations des extensions
Interprétation de la heatmap de corrélation



Utilisation des graphiques et calcul d'indices