

Optimisez la gestion et nettoyez les données du stock d'une boutique

Samuel OSENAT
Business Analyst
[Date de la présentation]

Analyses Exploratoires des Données

- *Dans ce projet on retrouve 3 Datasets : ERP, Web, Liaison*
- *Caractéristiques :*

ERP	Web	Liaison
825 observations 6 colonnes Pas de doublons	1513 observations 29 colonnes Beaucoup de doublons	825 observations 2 colonnes Pas de doublons

Le dataset liaison comporte un product_id et un ID Web ce qui permet de faire le rapprochement entre ERP et WEB. Dans liaison 91 occurrences ne correspondent pas entre les id product et web.

Analyses Exploratoires des Données

- *Traitement réalisés*
 - Nettoyages des données :
 - suppression des lignes où *post_type* était “attachment” car cela générait des doublons dans Web.
 - Conversion des lignes où le stock était négatif
 - Features engineering :
 - Sélection des colonnes utiles dans les deux dataframe
 - création de variables catégorielles dans *ERP* pour faire des vérifications (sur le stock par exemple)
- *Remarques éventuelles, pièges ou difficultés rencontrées*

Pas de difficultés spéciales rencontrées, hormis le fait que je n'ai pas ciblé de suite la dualité du post_type qui entraînait la présence de doublon.

Failli supprimer prix négatif aussi alors que ça peut avoir un sens métier.

Fusion ou consolidations des données

- *Choix des attributs*

Les attributs ont été sélectionnés sur la base d'une pertinence métier pour les analyses

- *Clés utilisées*

La fusion a été effectuée sur product_id entre erp et liaison puis entre id_web et sku

- *Vigilances particulières au cours du traitements*

Vigilance portée sur l'apparition de doublons suite aux jointures, et sur les lignes sans correspondances.

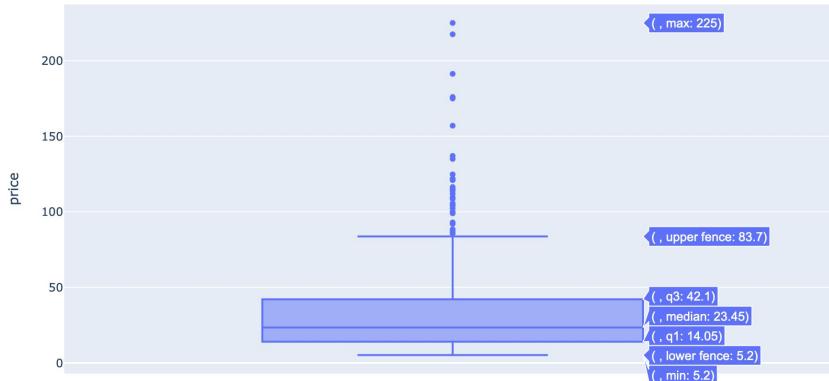
Les lignes pour lesquelles le SKU était nul ont été supprimées.

- *Difficultés ou pièges rencontrés*

Difficulté à trouver quel était la provenance des doublons

Analyses univariées du prix

- Méthodes statistiques employés :
Boite à moustache, méthode interquartile, z-score
Le seuil haut de prix identifié est de 84€ via la méthode interquartile
- Graphique avec commentaire des résultats



Ici, on voit bien que le seuil est à 84 euros

Analyses univariées du prix

- *Limites éventuelles de l'analyse*

La méthode du z-score nous donne un autre résultat sur le seuil de prix.

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

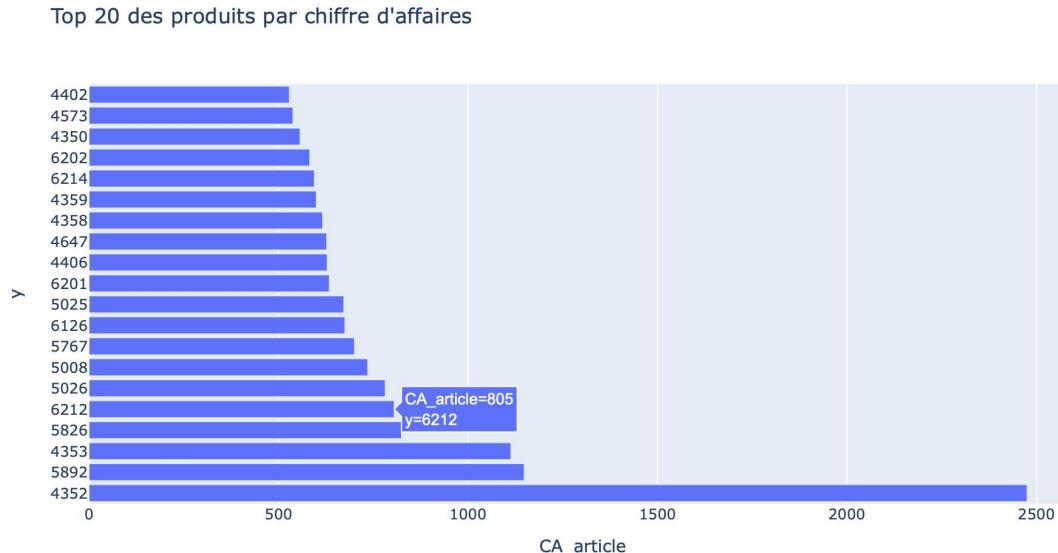
- *Méthodes statistiques employés*

Sommes, sommes cumulées 80/20, rotation des stocks sont les méthodes statistiques utilisées.

Ces méthodes nous permettent de calculer le CA total, le CA par article, les stocks par articles etc.

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

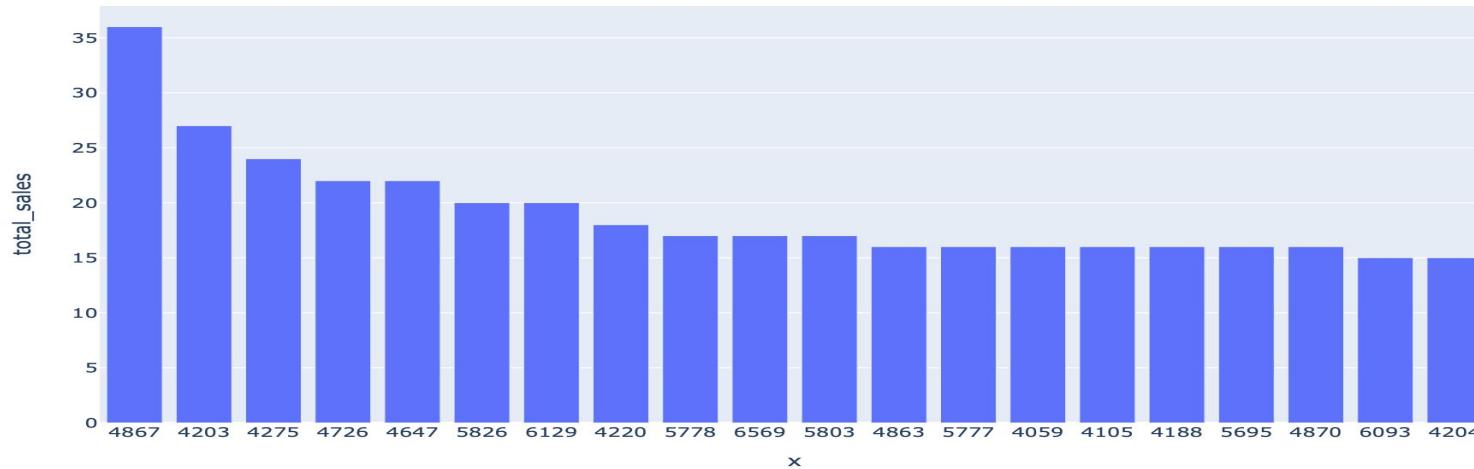
En calculant au préalable le CA par article on peut faire le top 20 des produits.



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

Analyse des ventes en quantité.

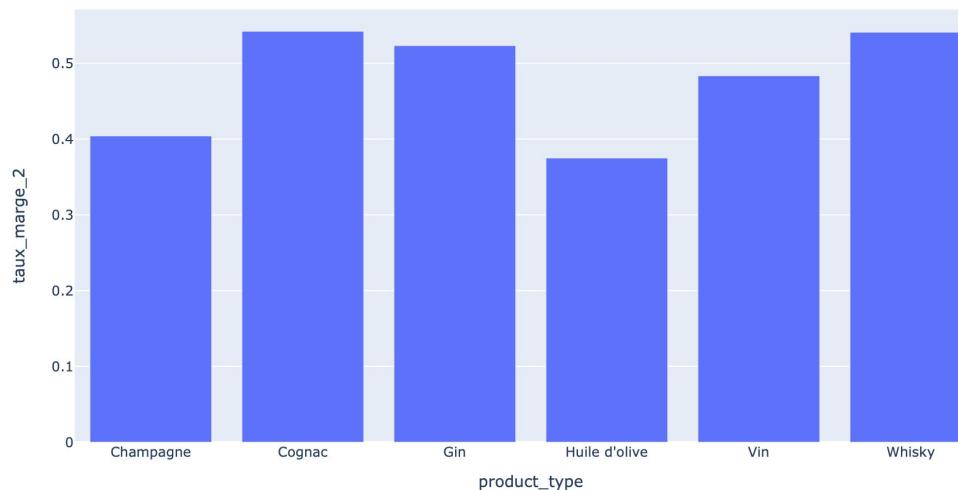
Les produits qui se vendent le plus en quantité ne sont pas forcément ceux qui génèrent le plus gros CA



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

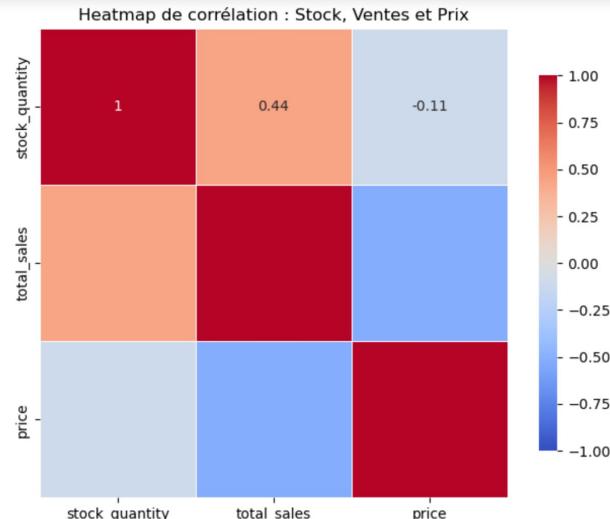
Analyse des taux de marge

Les produits comme le Cognac, le Gin, Whisky ainsi que le Vin ont un fort taux de marge. Cela indique leur positionnement sur le marché (haut de gamme).



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

Analyse des corrélations



- Relation positive entre quantités en stock et quantités vendues -> plus on a, plus on peut vendre
- Relation négative entre prix et quantités vendues -> prix bas entraîne plus de vente
- relation légèrement négative entre quantité en stock et prix

Actions pour la suite

- Pousser certaines analyses, quel % de produit représente quel % du CA ?
- Automatiser pipeline, fichier à la main pas bonne pratique
- Outil dataviz (Power BI, Tableau, Streamlit), potentielles alertes sur les prix
- voir anomalies du début et voir comment corriger avec équioies métiers et ERP

Point sur les compétences apprises

- *A l'aise avec python*
- *Qu'est-ce que vous avez trouvé le plus difficile ? Analyses univariées*
- *Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ? Prise de recul sur le projet, et ne pas être impatient.*