



PROJET 6

Jean-robert Vives

Analyses Exploratoires des Données
pour l'entité BOTTLENECK



Contexte

Analyste (nouvellement embauché) dans l'entreprise BottleNeck, un marchand de vin prestigieux. Il m'est demandé d'analyser 3 fichiers le premier provenant de l'exportation d'un CRM, le second d'un site internet et le troisième par un stagiaire

Objectifs

L'objectif est de transformer des données éparpillées et erronées en informations exploitables pour aider l'entreprise à mieux comprendre son activité et à prendre des décisions plus éclairées.



3 fichiers à disposition

- **erp.xlsx** (extraction de l'ERP)
- **web.xlsx** (extraction du site web)
- **liaison.xlsx** (une table de liaison reliant les références de l'ERP et du site web)

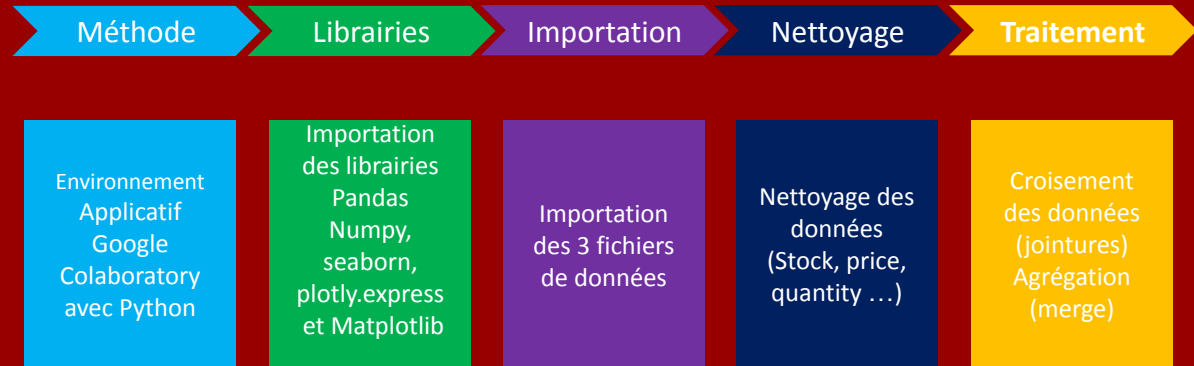
Environnement applicatif

GOOGLE COLLABORATORY avec Python



PHASE 1

Process : Nettoyage des données





PHASE 1

Nettoyage des données : les erreurs rencontrées

erp.xlsx



"Stock_status" incohérent
avec la valeur "stock_quantity"

web.xlsx



Des valeurs "sku"
absentes, incompréhensible
pour une exportation

liaison.xlsx



Des valeurs "id_web"
de type "objet"



Valeurs négatives
dans la colonne "price"



Valeurs en doublon
choix de garder "post-type"
égal "product"



"Sku" de web = "product_id" de erp
étrange pour la jonction car sku
devrait être aléatoire



"Stock_quantity" négatif
pour certaines valeurs



Des valeurs "sku"
de type "objet"



PHASE 2

Etude du CA

CA 143 769,20 €



PHASE 2

Analyse des ventes : 80/20



80 %

du chiffre d'affaire

=



434

articles
sur 714

61 %

du catalogue

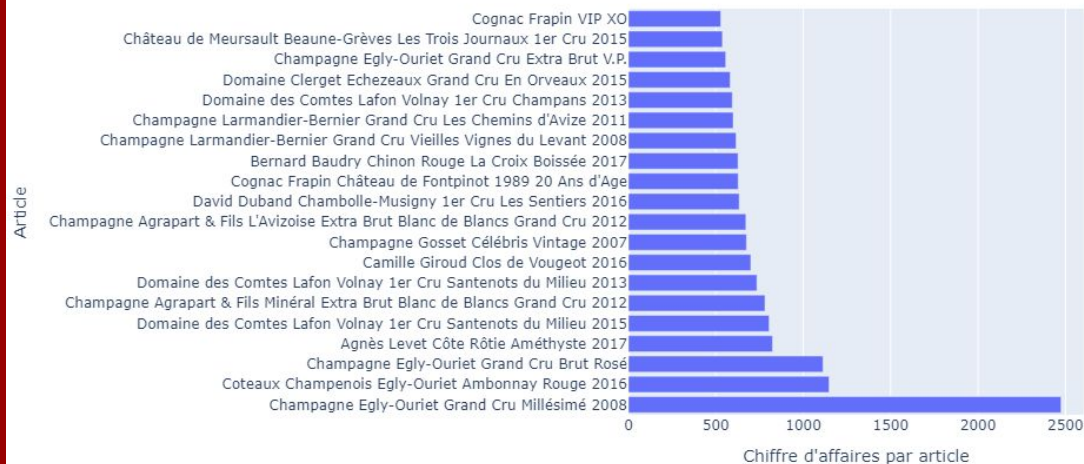


PHASE 2

Etude du CA

Les bouteilles qui génèrent le meilleur CA

Top 20 des articles qui ont généré le CA le plus important



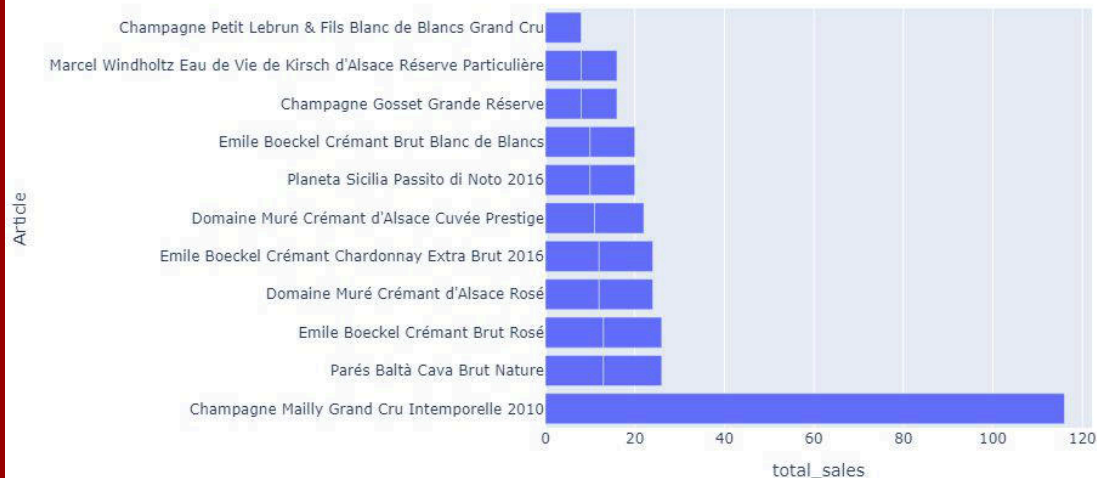


PHASE 2

Etude du CA

Les bouteilles les plus vendues

Top 20 des articles les plus vendus





PHASE 2

Etude des valeurs aberrantes : Les méthodes

Méthode 1
**intervalle
interquartile**

Méthode 2
Z-SCORE

Sachant que la méthode interquartile dépend d'un coef avec une convention qui peut varier selon les produits, je préfère utiliser la méthode 2 qui me parait plus lisible en fonction de mes connaissances !



PHASE 2

Les valeurs aberrantes : Le résultat
Méthode 2 - Z-SCORE

29 Articles
"ABERRANTS"

+87,53 €

12,2 % du CA

Suggestion : STOCK aberrant ➡

Nouvelles références non aberrantes : ➡ STOCK



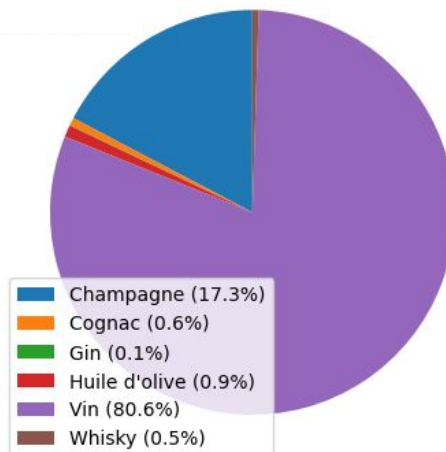
PHASE 2

Etude du stock

16 740



Répartition des bouteilles par type





PHASE 2

Etude du stock

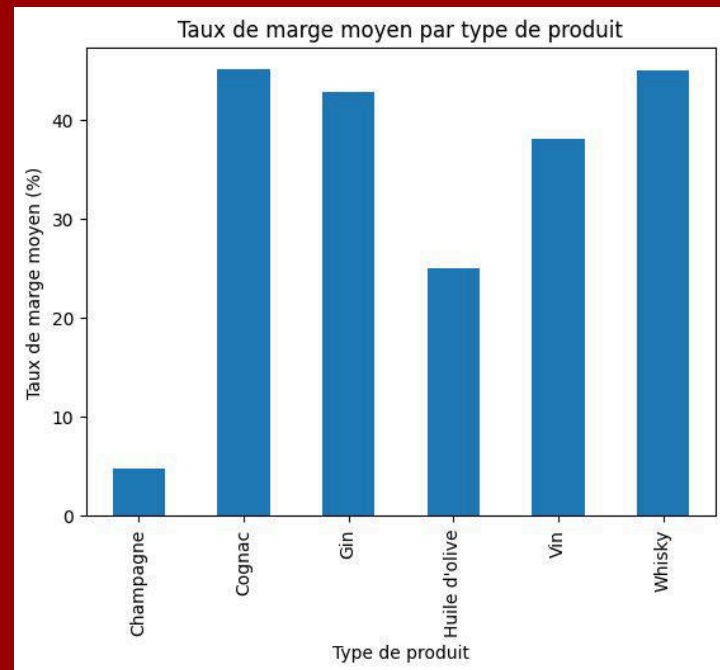
VALEUR STOCK
277 328,07 €



PHASE 2

Etude du stock

Taux de marge moyen par type de produit

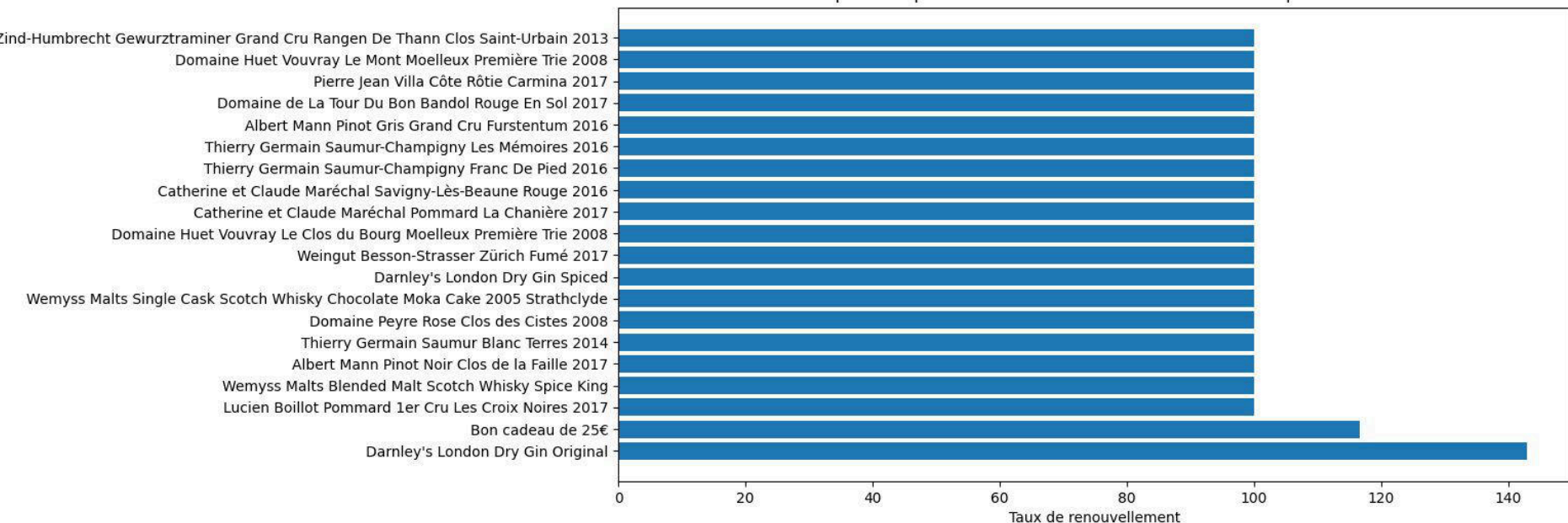




PHASE 2

Rotation des stocks

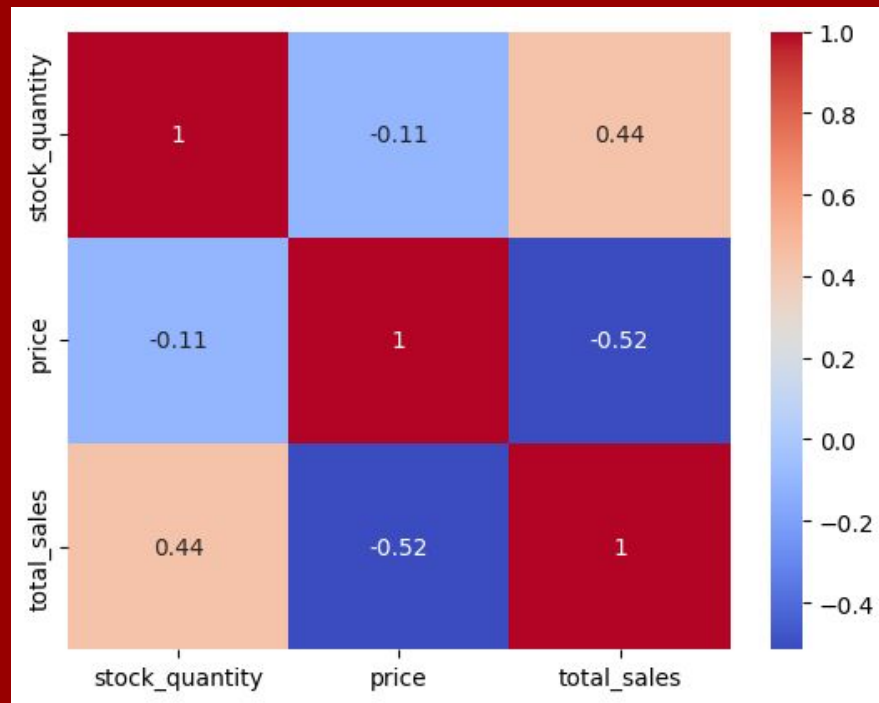
Top 20 des produits avec le taux de renouvellement le plus élevé





PHASE 2

Analyse des corrélation





PHASE 2

Analyse des corrélation

0,44

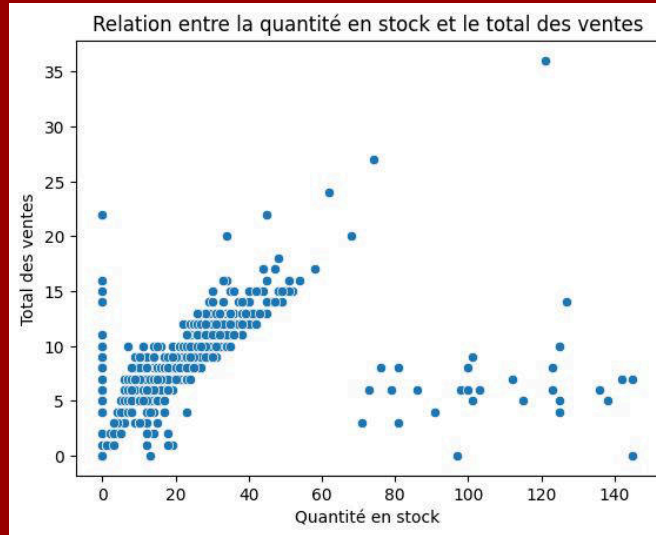
Stock_quantity ↗

total_sales ↗



PHASE 2

Etude du stock



On remarque des produits aberrants qui s'éloignent de la corrélation, beaucoup de stock mais les ventes ne sont pas en adéquation, on pourra étudier ces produits ultérieurement et voir si ils ne sont pas en fin de vie



PHASE 3

Conseils et suggestions futur ERP

- **Génération auto** de l'indication “outofstock” ou “instock” à la création du stock
- Empêcher la décrémentation du stock (stock négatif) si stock 0 sur CRM (inventaire ou process de mise à jour des stocks à la réception de la marchandise)
- Définir une règle de calcul entre “purchase price” et “price” (coef fixe par catégorie de produit)
- Utiliser un champ commun “id” sur l'ERP et transformer “sku” sur le web en “id” sur chaque produit ce qui assurera une jointure rapide et plus efficace avec le fichier web et éviter le fichier liaison