

Optimisez la gestion des données d'une
boutique avec R ou Python

Rouhling Jules
Cursus Data Analyst
19/01/2023

Analyses Exploratoires des Données

3 bases de données

ERP

prix achat

prix vente

ID produit

stock

onsale_web

825 lignes

WEB

SKU

ventes

type produit

1513 lignes

LIAISON

SKU

ID produit

825 lignes

Analyses Exploratoires des Données

Nettoyage des données ERP

product_id	onsale_web	price	stock_quantity	stock_status	purchase_price
------------	------------	-------	----------------	--------------	----------------

Création nouvelle colonne stock_status

2 erreurs trouvées

Vérification quantités négatives

2 données

→ corrigées à 0

Vérification des doublons ID produit

RAS

Vérification prix achat négatifs

RAS

Vérification onsale_web

1 = en vente / 0 = pas en vente

Vérification prix vente négatifs

3 données négatives

→ vérification de la
cohérence avec prix
d'achat

→ simple problème de
signe

Analyses Exploratoires des Données

Nettoyage des données WEB

1- conservation des colonnes 'sku', 'total_sales', 'tax_status', 'product_type'

2- vérification des 'sku' atypiques

[nan, '13127-1', 'bon-cadeau-25-euros']

3- gestion des 'sku' atypiques

suppression des bons d'achat + suppression des valeurs nulles

4- gestion du 'tax_status'

chaque produit a deux lignes identiques (une taxable et une nulle) → suppression des lignes nulles pour éviter doublons et suppression colonne 'tax_status'

Fusion ou consolidations des données

DF_MERGE = fusion de LIAISON et ERP

825 observations correspondantes sans doublon dans les deux tables

→ Fusion sur clé 'product_ID'

DATA = fusion de **DF_MERGE** et WEB

certaines ID produit sans SKU correspondant

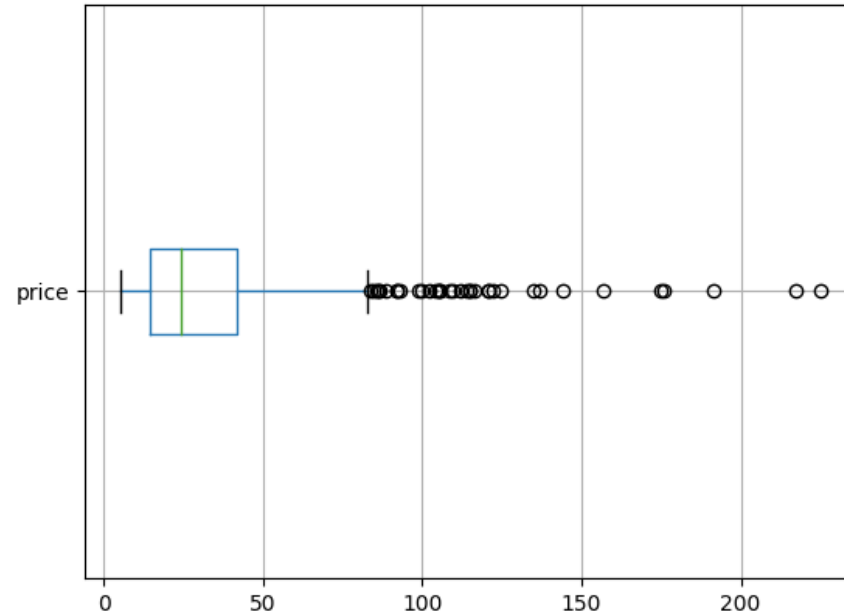
→ conservation toutes valeurs même celle sans infos dans WEB

→ fusion sur clé 'sku'

→ absence infos ventes sur 112 produits non vendus (825 – 713)

Analyses univariées du prix

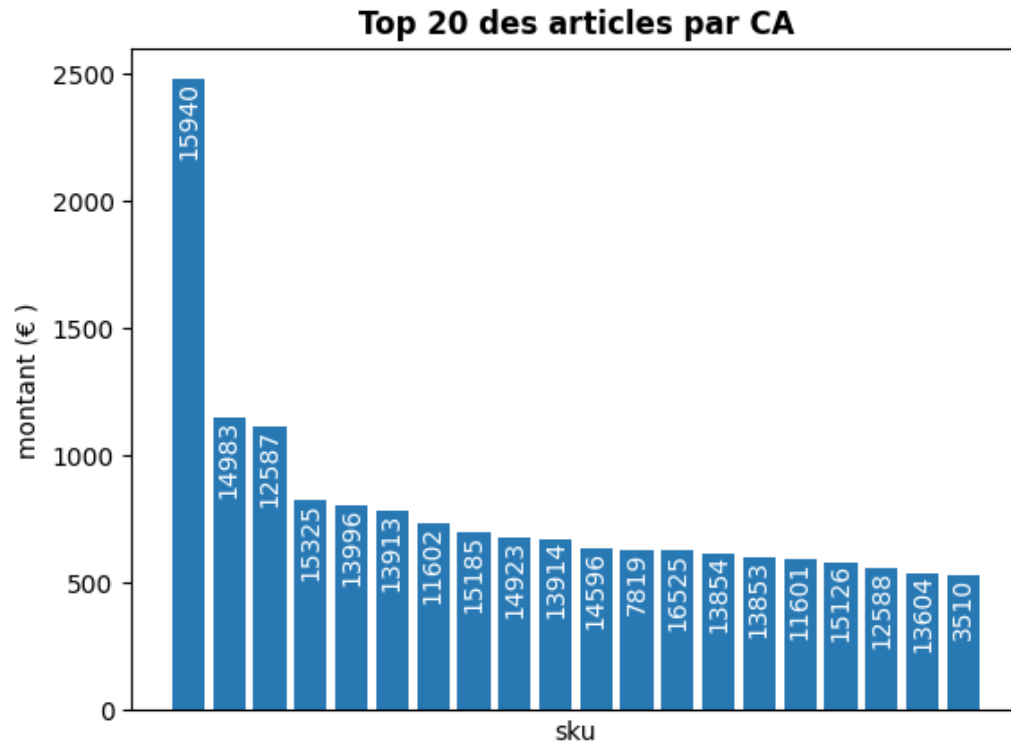
DATA boîte à moustache de la variable 'prix'



Analyses complémentaires

Chiffre d'affaires

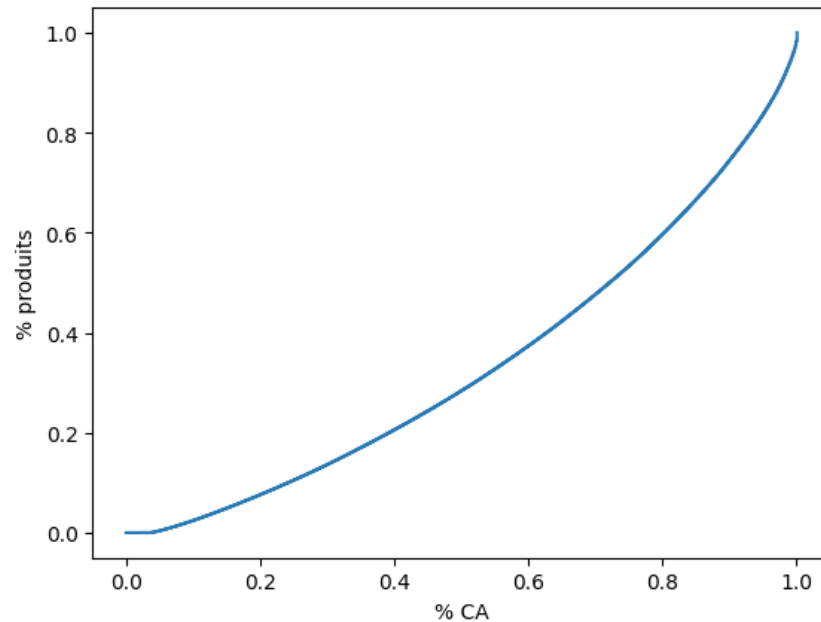
CA total octobre
EUR 143,505



Analyses complémentaires

Chiffre d'affaires

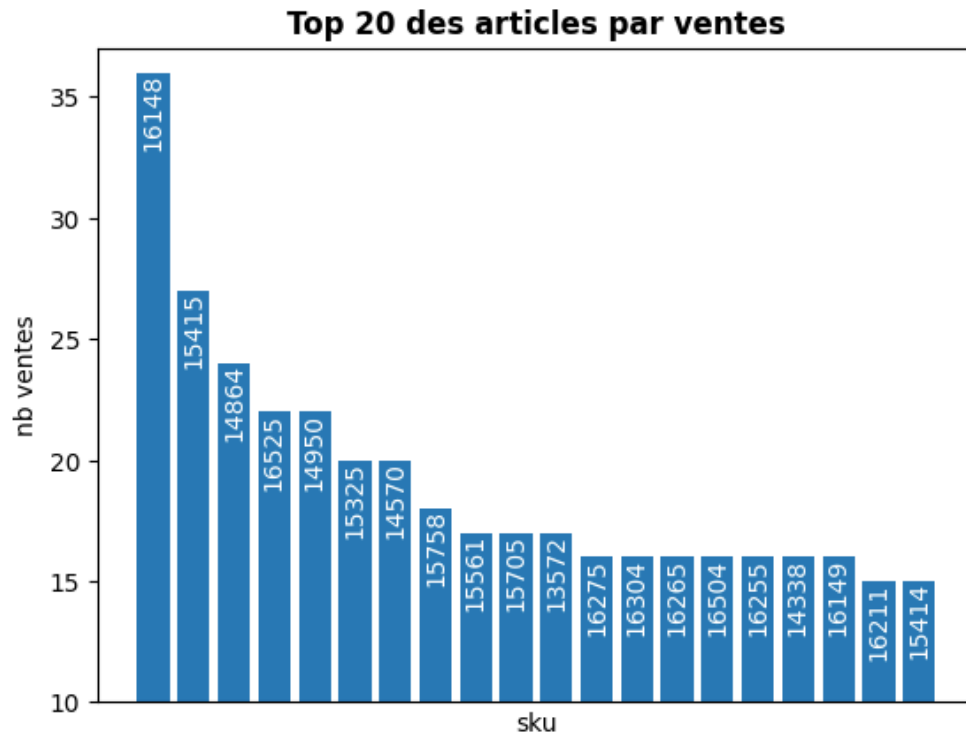
DATA courbe de Pareto % produits / % CA



Analyses complémentaires

Quantités

DATA
nb ventes octobre
5,744



Analyses complémentaires Stocks

DATA

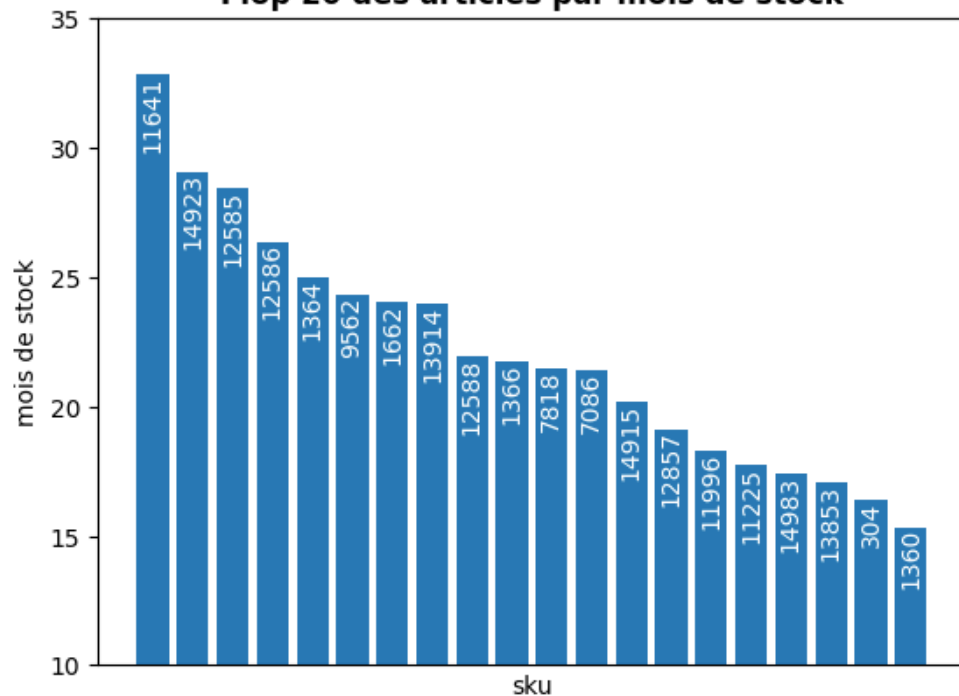
valorisation stocks 31
octobre

EUR 298,627

stocks (unités)

17,822

Flop 20 des articles par mois de stock



Analyses complémentaires

Taux de marge

DATA

tous produits

tx de marge min

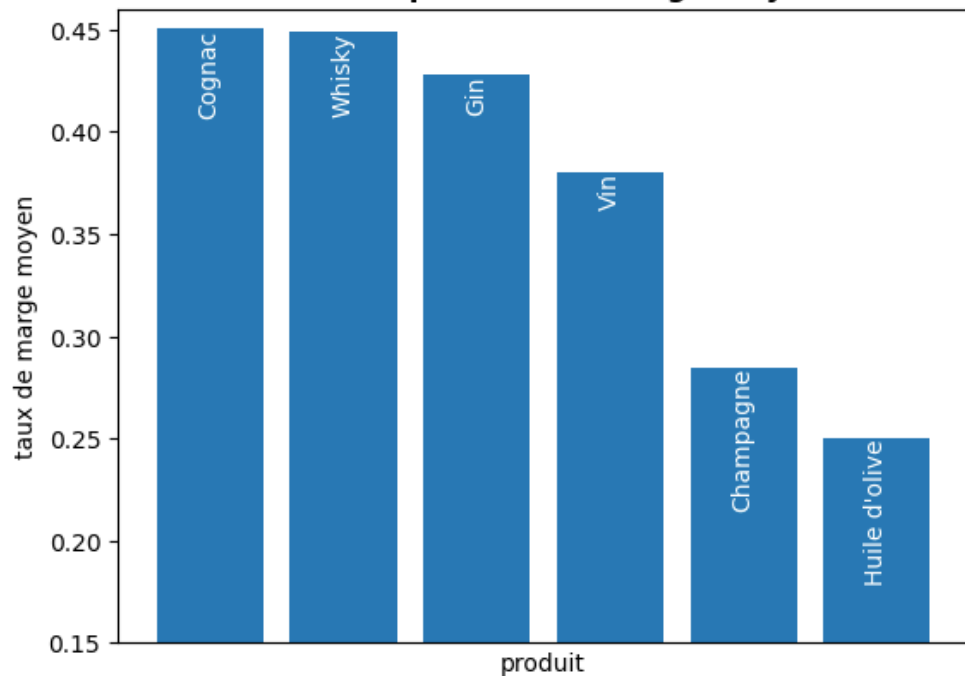
0.23

tx de marge max

0.48

(4 valeur négatives
supprimées)

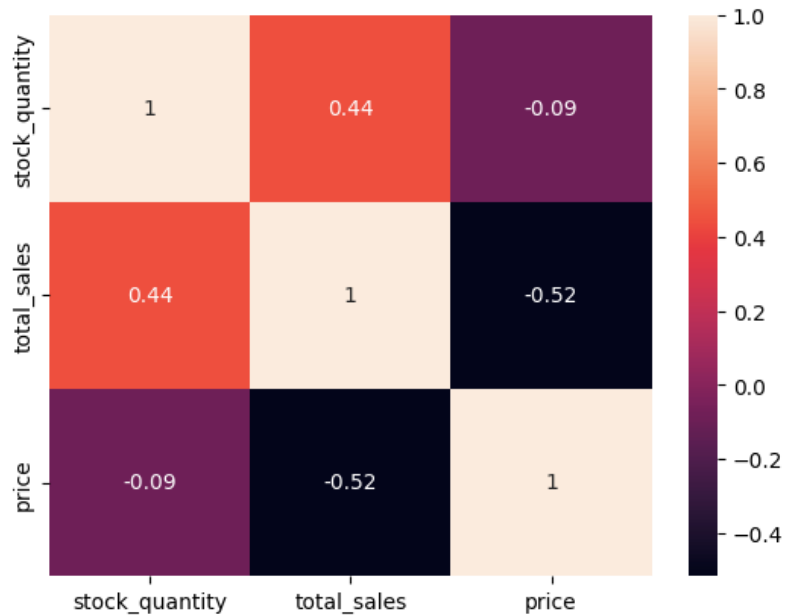
Produits par taux de marge moyen



Analyses complémentaires

Corrélations

DATA tableau Seaborn des corrélations entre prix, ventes et stocks



Actions pour la suite

- 1- si possible associer un 'sku' (données ventes du mois) à un 'product_id' (caractéristiques produits) dans la table LIAISON**
- 2- doit-on supprimer les produits à faible taux de marge (<0.3) ?**
- 3- il faut regarder de près les produits avec une rotation des stocks très élevée (> 12 mois?)**