



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

Jérémie QUERET

Les données initiales : 3 fichiers

ERP

1



825 Observations
6 Colonnes

LIAISON

2



825 Observations
2 Colonnes

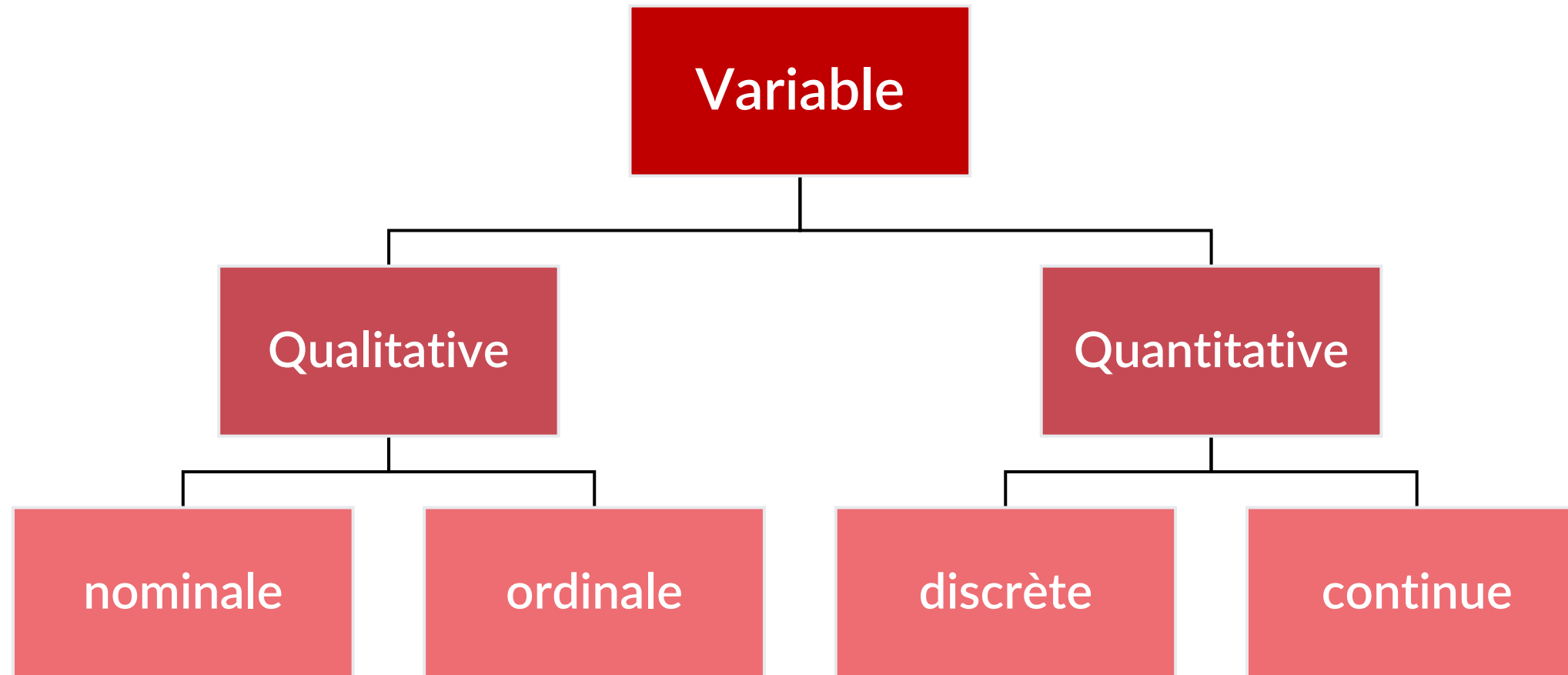
WEB

3

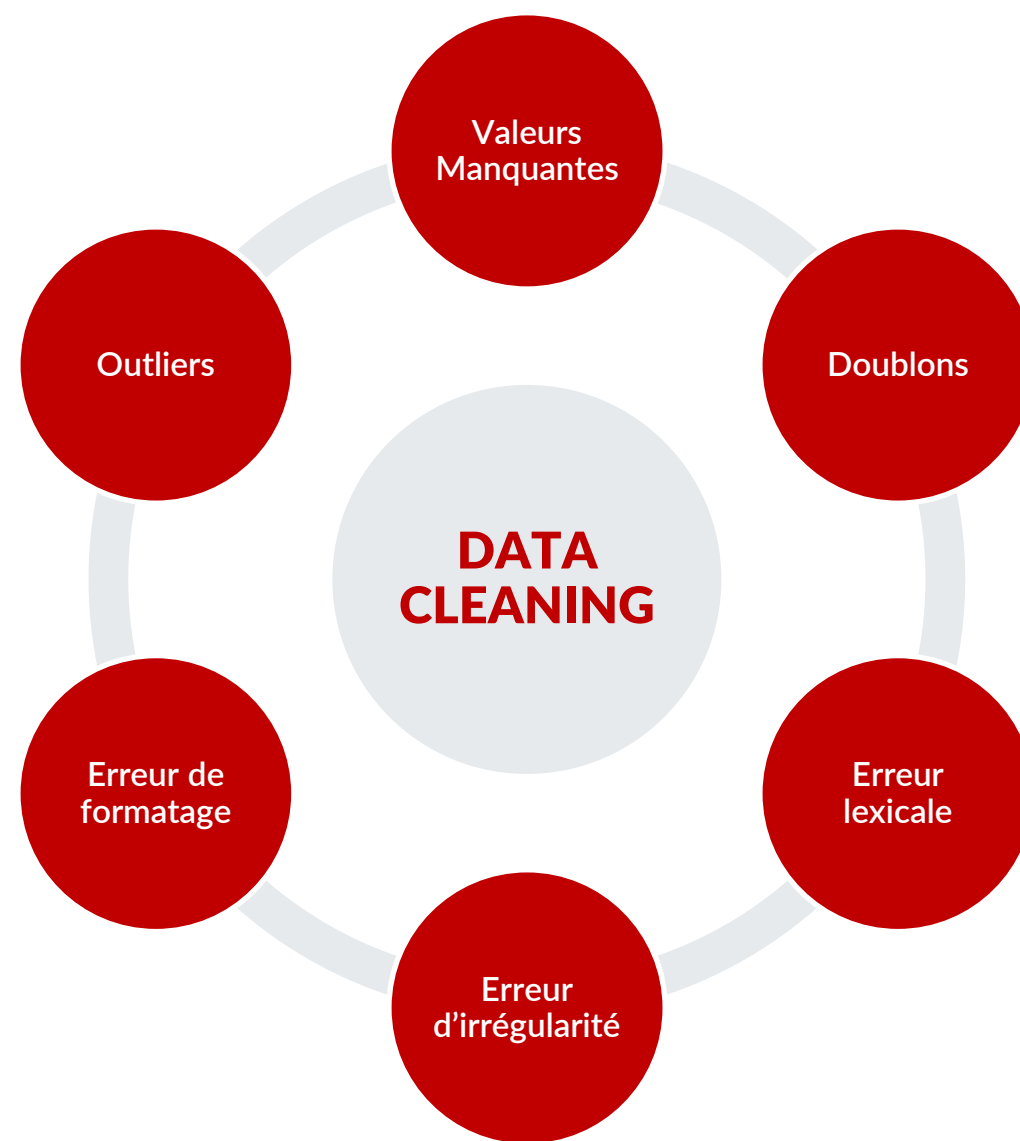


1513 Observations
29 Colonnes

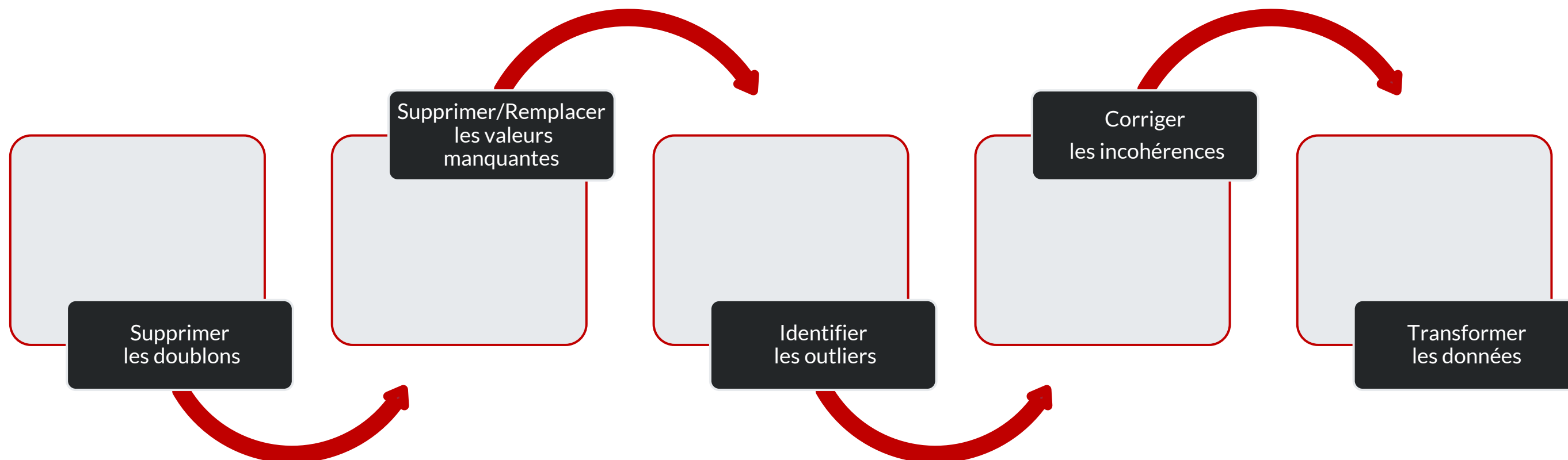
Les types de variables



Nettoyer les données



Nettoyer les données : le traitement



Consolider les données : outer merge

product_id	X2
A	1
B	2
C	3

df_erp



product_id	X3
A	T
B	F
D	T

df_liaison



product_id	X2	X3
A	1	T
B	2	F
C	3	NaN
D	NaN	T

df_merge

```
df_merge = pd.merge(df_erp, df_liaison, on='product_id',  
                    how = 'outer', indicator=True)
```

Conserver toutes les valeurs,
toutes les lignes

Consolider les données : outer merge

sku	X2
A	1
B	2
C	3

df_merge



sku	X3
A	T
B	F
D	T

df_web



sku	X2	X3
A	1	T
B	2	F
C	3	NaN
D	NaN	T

df_final

```
df_final = pd.merge(df_merge, df_web, on='sku', how='outer',  
                    indicator=True)
```

Conserver toutes les valeurs,
toutes les lignes

Le dataframe final

df_final

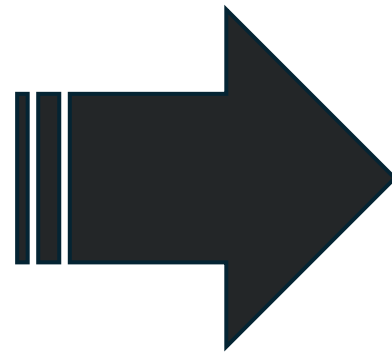
#



825 Observations
30 Colonnes

Feature Engineering : création de variables

11
Variables



df_final

#

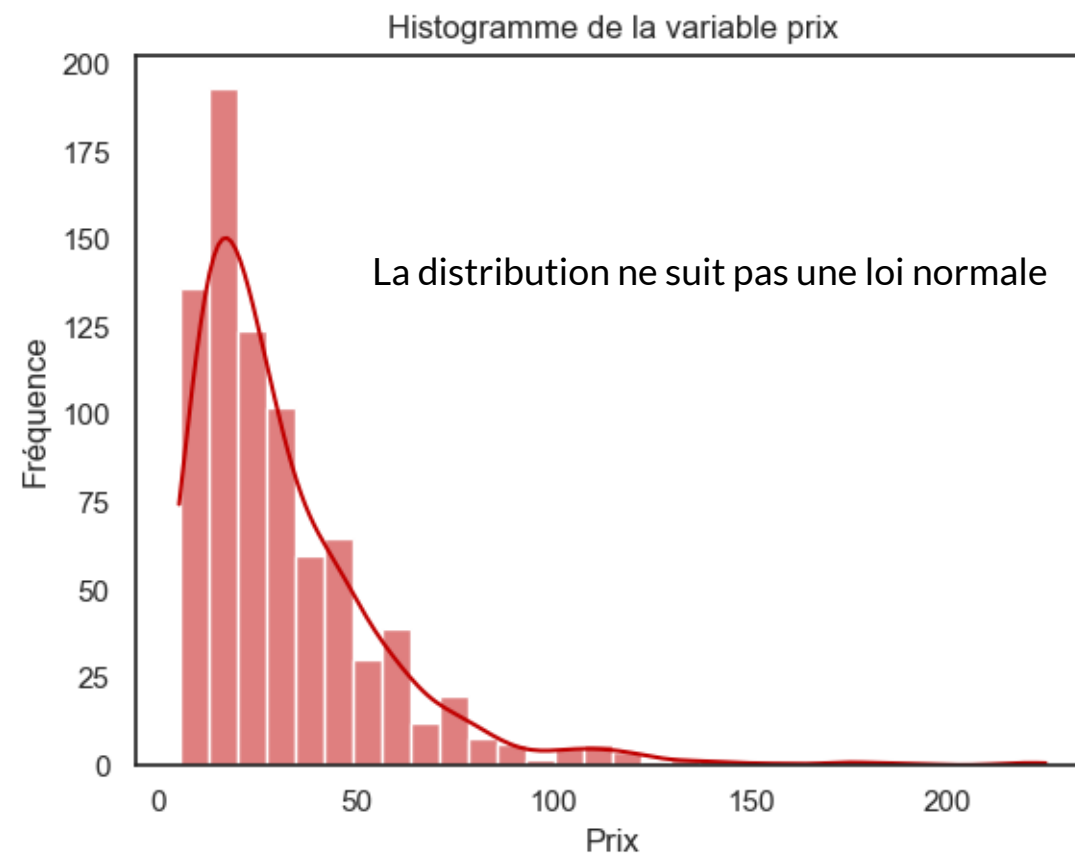
- CA par article
- CA en %
- CA en % cumulé
- Total sales en %
- Total sales en % cumulé
- Rotation des stocks
- Valorisation de stock
- Prix HT
- Marge commerciale
- Taux de marge
- Taux de marque



Analyses des données



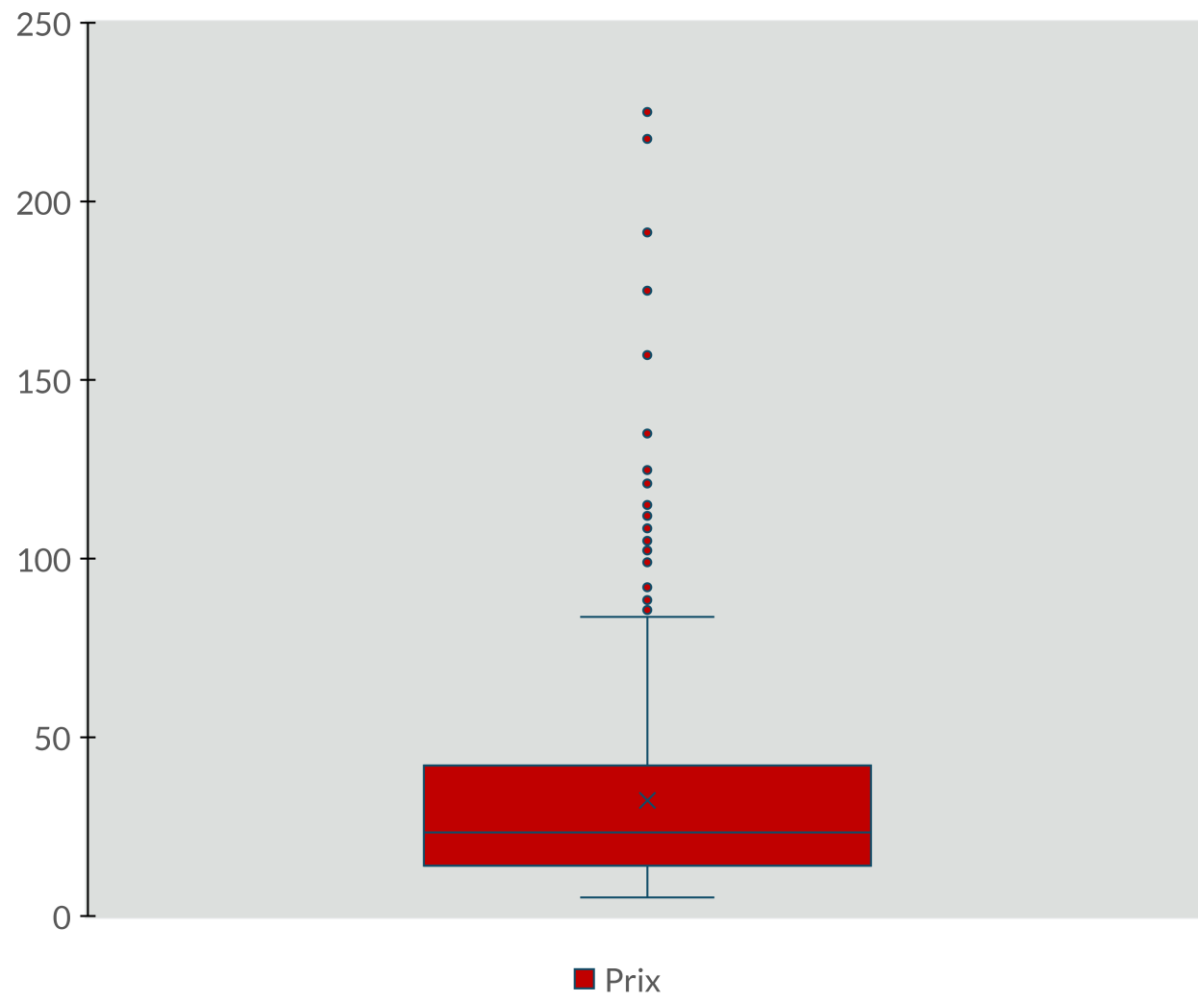
Analyse de la variable prix : distribution



Caractéristiques d'une distribution asymétrique à droite :

- La moyenne est généralement supérieure à la médiane.
- La queue sur le côté droit est plus longue et contient souvent des valeurs aberrantes.
- La majorité des points de données se trouvent sur le côté gauche.

IQR : 36 outliers représentent 4,36% du catalogue.



Max

225

Moyenne

32,35

Médiane

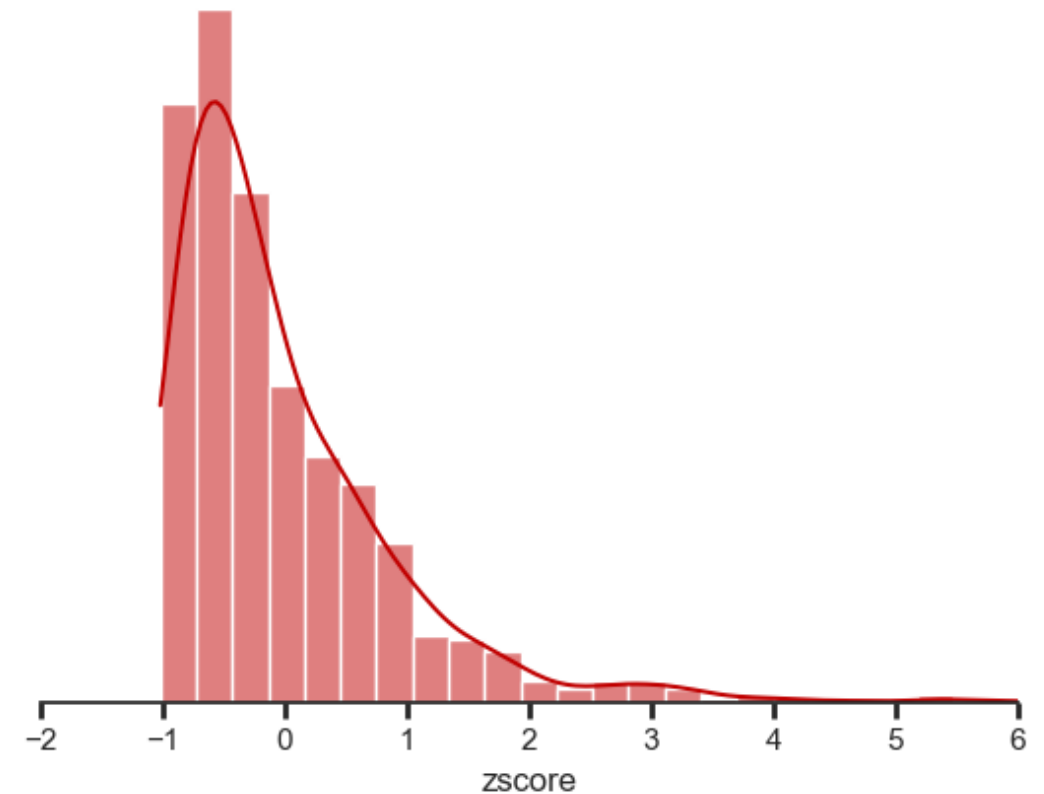
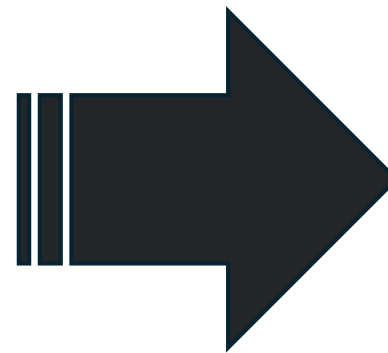
24,4

Min

5,2

Zscore : 17 valeurs aberrantes au seuil supérieur à 3

Zscore > 3
17 Outliers



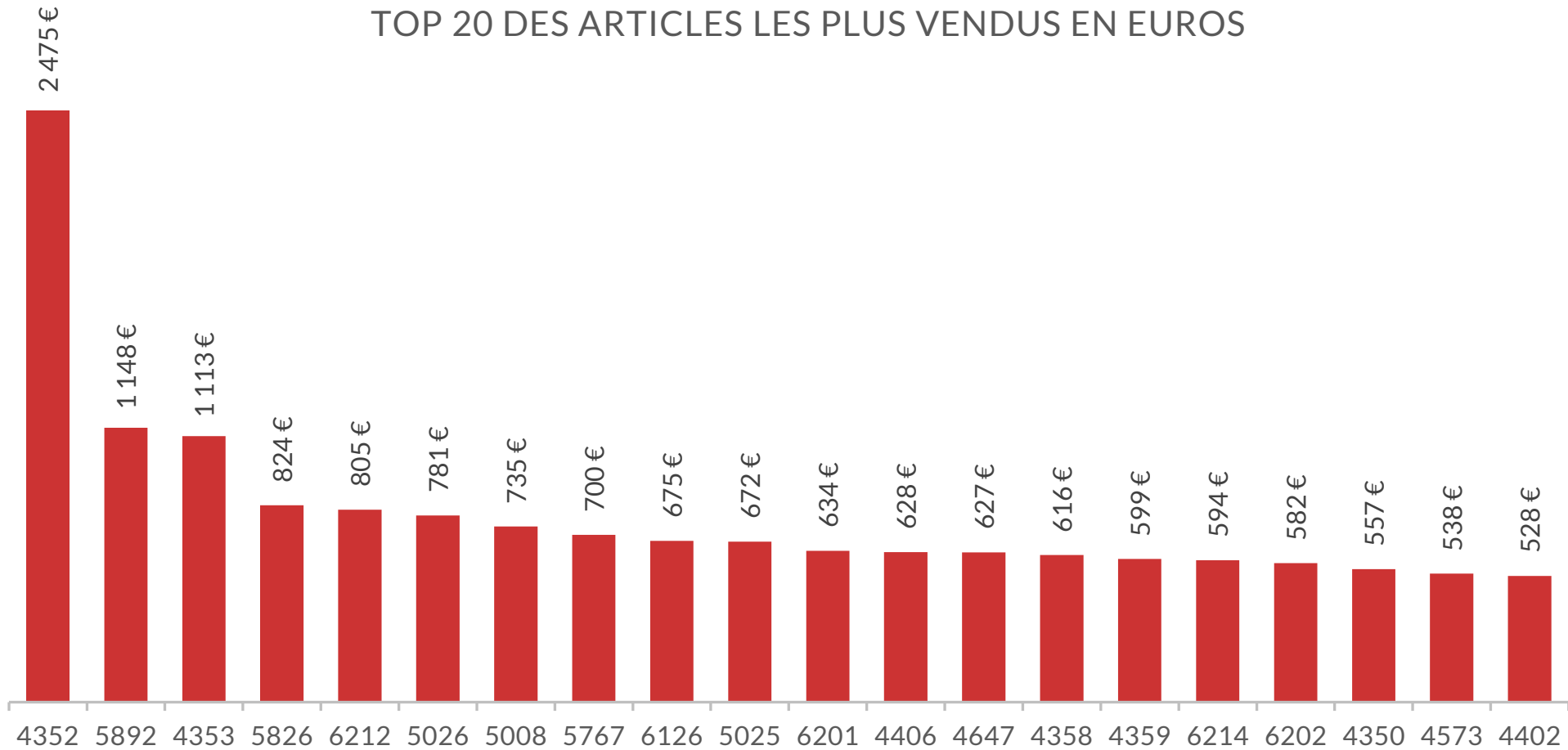


Analyse du chiffre d'affaires

Chiffre d’Affaires
143 680 €

Nb de
commandes
716

TOP 20 DES ARTICLES LES PLUS VENDUS EN EUROS



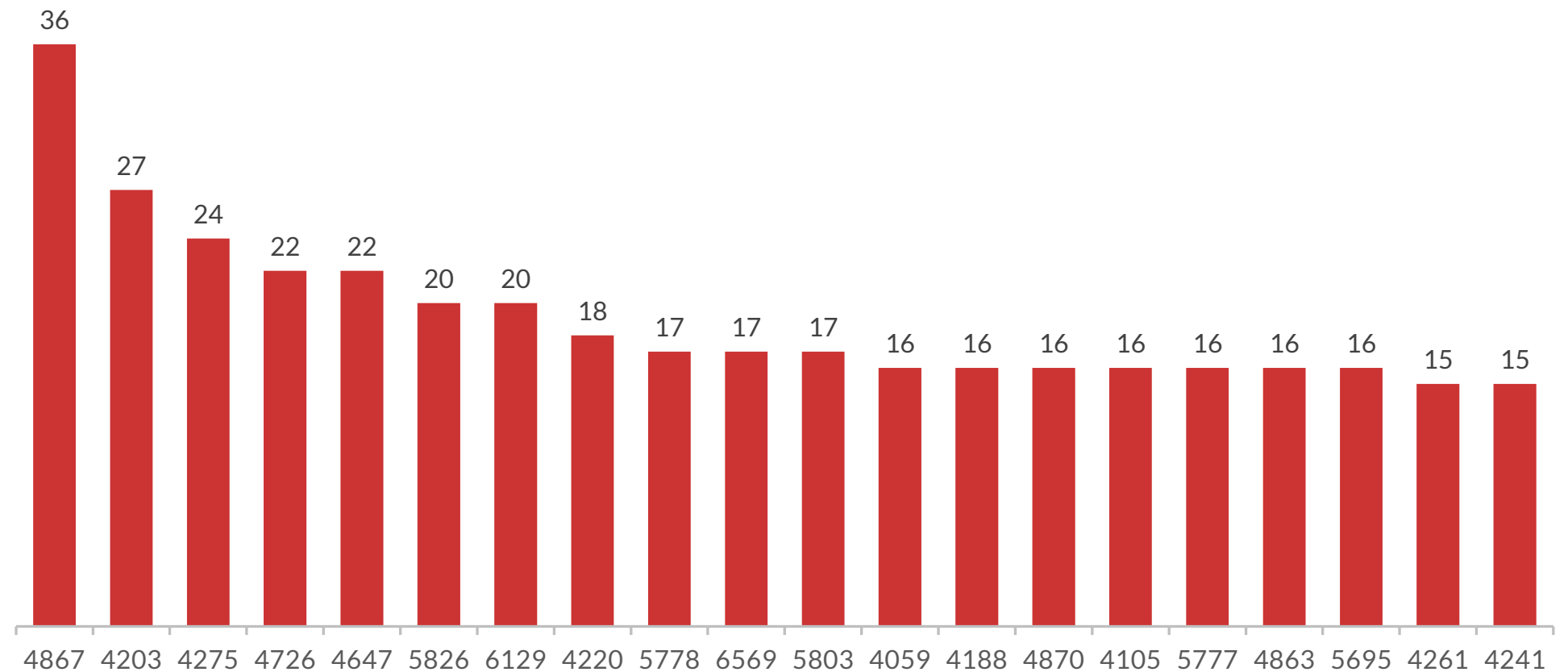


80₂₀ en quantité d'articles vendus

TOP 20 DES ARTICLES LES PLUS VENDUS EN QUANTITÉ

441
Articles

53,5%
du catalogue





Analyse des stocks : 20 articles en rupture

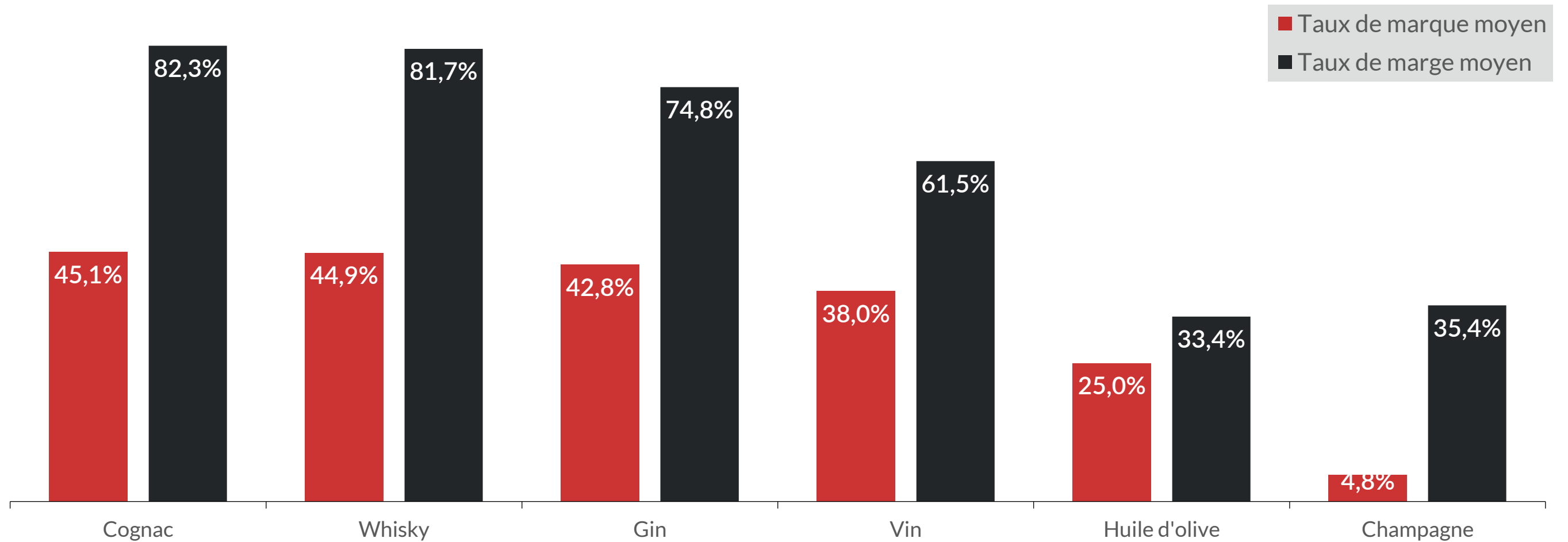
LES 20 ARTICLES EN RISQUE DE RUPTURE DE STOCK

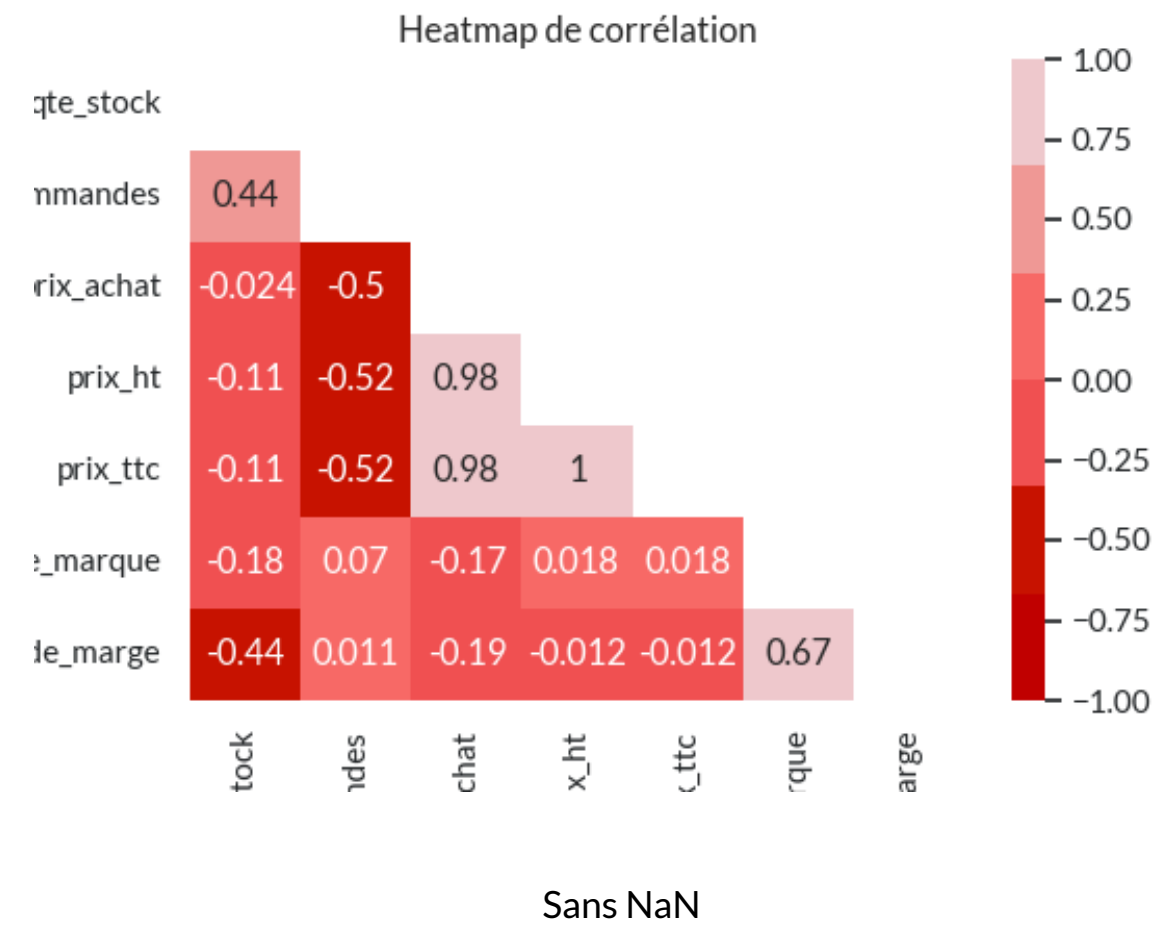
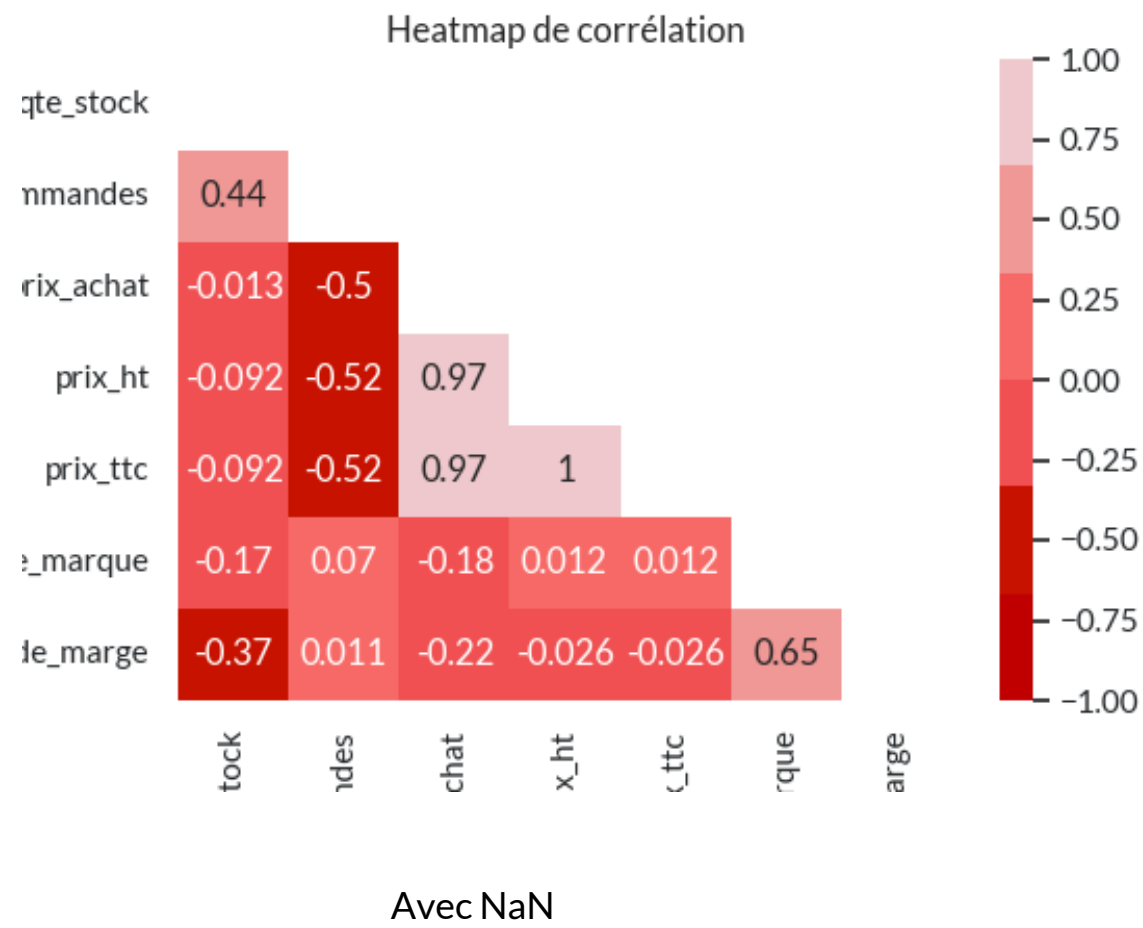
Qté stock
17 822

Valorisation
298 627,7 €

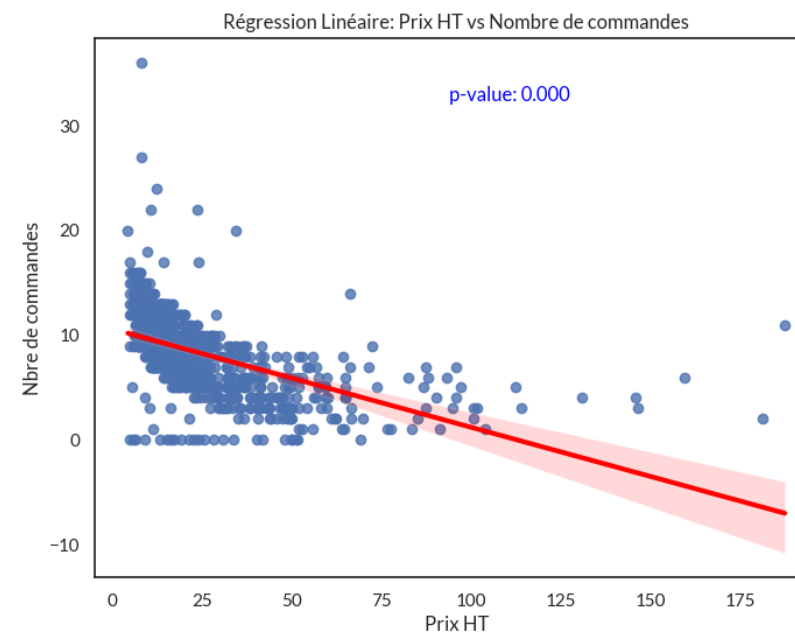
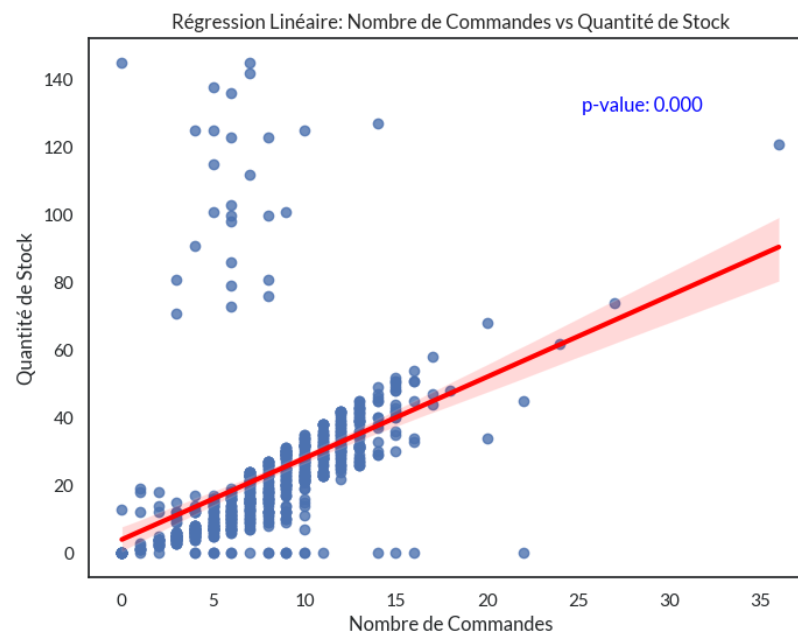


Analyse de la marge





Analyse de l'activité





Recommandations

1. Les clés product_id et sku doivent comporter aucunes données manquantes (contrainte NOT NULL dans MySQL)
2. Les variables : qté_stock ,nbre_commandes, prix_achat, prix_ht, prix_ttc, taux_de_marque, taux_de_marge doivent également être toutes renseignées , ainsi une contrainte NOT NULL est à appliquer dans MySQL .
3. 91 id_web sont manquants dans le fichier liaison. Ce qui nécessite de supprimer les id qui ne correspondent à aucun article si ceux-ci ne sont plus commercialisés.
4. Configurer des règles au sein de la plateforme ecommerce afin d'éviter des erreurs de prix . (prix ttc < au prix d'achat)
5. Normaliser la table web et insérer entre autres, une table promotion avec ses attributs dédiés.
6. Améliorer la gestion des stocks avec une solution de gestion de stock en temps réel.