

Analyse_Vins_Spiritueux

September 5, 2025

Analyse et Consolidation des Données - Vins & Spiritueux

0.1 Données

0.1.1 Données issues de l'ERP (gestion interne)

| Colonne | Description |
|-----------------------------|--|
| <code>product_id</code> | Identifiant unique du produit dans le système ERP |
| <code>onsale_web</code> | Indique si le produit est en vente sur le site web (1 = oui, 0 = non) |
| <code>price</code> | Prix de vente du produit (affiché sur le site web) |
| <code>stock_quantity</code> | Quantité de stock disponible pour le produit |
| <code>stock_status</code> | État du stock (<code>instock</code> , <code>outofstock</code> , etc.) |
| <code>purchase_price</code> | Prix d'achat du produit (coût pour l'entreprise) |

0.1.2 Données issues du site web

| Colonne | Description |
|-----------------------------|--|
| <code>sku</code> | Code produit unique utilisé sur le site web (identifiant de référence) |
| <code>virtual</code> | Produit virtuel (non physique), booléen (1 ou 0) |
| <code>downloadable</code> | Produit téléchargeable (ex. : PDF, logiciel) |
| <code>rating_count</code> | Nombre total d'avis/notes reçues par le produit |
| <code>average_rating</code> | Note moyenne attribuée au produit (généralement sur 5) |
| <code>total_sales</code> | Nombre total d'unités vendues |
| <code>tax_status</code> | Statut fiscal du produit (<code>taxable</code> , <code>none</code> , etc.) |
| <code>tax_class</code> | Classe de taxe applicable (si renseignée) |
| <code>post_author</code> | ID de l'auteur (utilisateur qui a créé/modifié le produit) |
| <code>post_date</code> | Date de création de la fiche produit |
| <code>post_date_gmt</code> | Date de création (heure GMT) |
| <code>post_content</code> | Description longue du produit |
| <code>product_type</code> | Type de produit (<code>simple</code> , <code>variable</code> , etc.) |
| <code>post_title</code> | Titre du produit |
| <code>post_excerpt</code> | Description courte (extrait) |
| <code>post_status</code> | Statut de publication (<code>publish</code> , <code>draft</code> , etc.) |
| <code>comment_status</code> | Autorisation ou non des commentaires (<code>open</code> , <code>closed</code>) |
| <code>ping_status</code> | Notifications externes (peu utilisé pour les produits) |

| Colonne | Description |
|-----------------------|---|
| post_password | Mot de passe d'accès (rarement utilisé) |
| post_name | Slug du produit (nom SEO-friendly utilisé dans l'URL) |
| post_modified | Date de dernière modification |
| post_modified_gmt | Idem, en heure GMT |
| post_content_filtered | Contenu nettoyé (souvent vide ou non utilisé) |
| post_parent | ID du parent (utilisé pour les produits enfants ou groupés) |
| guid | Lien unique généré automatiquement par WordPress |
| menu_order | Ordre d'affichage du produit dans les menus |
| post_type | Type d'objet (product , page , etc.) |
| post_mime_type | Type MIME du fichier joint (souvent vide) |
| comment_count | Nombre de commentaires ou d'avis laissés |

0.1.3 Données de liaison entre ERP et site web

| Colonne | Description |
|------------|--|
| id_web | Identifiant du produit dans la base web (souvent équivalent à sku) |
| product_id | Identifiant du produit dans la base ERP (utilisé pour la correspondance interne) |

```
[ ]:
```

```
[ ]:
```

0.2 création de fonctions pour identifier et afficher les outliers

```
[7]: def detection_des_outliers(df , col ) :

    Q1 = np.percentile(df[col],25 )

    Q3 = np.percentile(df[col],75)

    IQR = Q3 - Q1

    BI = Q1 - 1.5 * IQR
    BS = Q3 + 1.5 * IQR

    les_outliers = df[ (df[col] < BI ) | ( df[col] > BS ) ]

    print(f" dans la colonne {col} nous avons {len(les_outliers)} outliers" )

    fig = px.box(df , y = col , title=f" Boxplot de {col} avec les outliers " )
    fig.update_layout(height = 500 , width = 600)
    fig.show()
```

```

    return les_outliers

def Affichage_des_outliers(df, colonne):
    q1 = df[colonne].quantile(0.25)
    q3 = df[colonne].quantile(0.75)
    iqr = q3 - q1
    borne_inf = q1 - 1.5 * iqr
    borne_sup = q3 + 1.5 * iqr
    return (df[colonne] < borne_inf) | (df[colonne] > borne_sup)

```

[]:

[]:

Importation des librairies et chargement des fichiers

Importation des librairies

```

[13]: import pandas as pd
import numpy as np
import plotly.express as px
from scipy.stats import zscore

from plotly.subplots import make_subplots
import plotly.graph_objects as go
import matplotlib.pyplot as plt
import seaborn as sns

```

Chargement des fichiers

```

[15]: erp = pd.read_csv("C:/Users/lklk/OneDrive/Documents/cours MASTER/DATA SCIENCE/
↳examen/DataSet/erp.csv")
web = pd.read_csv("C:/Users/lklk/OneDrive/Documents/cours MASTER/DATA SCIENCE/
↳examen/DataSet/WEB.csv")
liaison = pd.read_csv("C:/Users/lklk/OneDrive/Documents/cours MASTER/DATA_
↳SCIENCE/examen/DataSet/liaison.csv")

```

Analyse exploratoire des fichiers

Phase 1 : Consolidation des données.

Analyse de la table ERP.

0.3 Verifiacion des dimensions

```

[22]: ligne,colonne = erp.shape

print(f" la table ERP contient {ligne} ligne(s) et {colonne} colonne(s) ")

```

la table ERP contient 825 ligne(s) et 6 colonne(s)

0.4 Vérification des valeurs nulles

```
[25]: erp.isnull().sum()
```

```
[25]: product_id      0
      onsale_web      0
      price          0
      stock_quantity  0
      stock_status    0
      purchase_price  0
      dtype: int64
```

On constate aisément que la table ERP ne dispose d'aucune valeur nulle ou manquante sur l'ensemble de ses colonnes

```
[ ]:
```

0.5 Verification des doublons

```
[30]: doublon_erp = erp.duplicated().sum()
      print(f" le table contient {doublon_erp} doublon(s) ")
```

```
le table contient 0 doublon(s)
```

```
[32]: erp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 825 entries, 0 to 824
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   product_id      825 non-null   int64
1   onsale_web      825 non-null   int64
2   price           825 non-null   float64
3   stock_quantity  825 non-null   int64
4   stock_status    825 non-null   object
5   purchase_price  825 non-null   float64
dtypes: float64(2), int64(3), object(1)
memory usage: 38.8+ KB
```

On peut constater que la table ERP dispose d'un typage normal pour chacune de ses colonnes

```
[35]: erp
```

```
[35]:   product_id  onsale_web  price  stock_quantity  stock_status \
0         3847           1   24.2             16      instock
1         3849           1   34.3             10      instock
2         3850           1   20.8              0  outofstock
3         4032           1   14.1             26      instock
4         4039           1   46.0              3  outofstock
```

```

..      ...      ...      ...      ...      ...
820      7203      0      45.0      30      instock
821      7204      0      45.0      9      instock
822      7247      1      54.8      6      instock
823      7329      0      26.5      14      instock
824      7338      1      16.3      40      instock

```

```

      purchase_price
0      12.88
1      17.54
2      10.64
3      6.92
4      23.77
..      ...
820      23.48
821      24.18
822      27.18
823      13.42
824      8.00

```

```
[825 rows x 6 columns]
```

```
[ ]:
```

Analyse de la table WEB.

0.6 Vérification des dimensions

```
[40]: ligne_W ,colonne_W = web.shape
      print(f" la table WEB contient {ligne_W} ligne(s) et {colonne_W} colone(s) ")
```

```
la table WEB contient 1513 ligne(s) et 29 colone(s)
```

```
[42]: web.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1513 entries, 0 to 1512
Data columns (total 29 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sku                    1428 non-null  object
1   virtual                1513 non-null  int64
2   downloadable           1513 non-null  int64
3   rating_count           1513 non-null  int64
4   average_rating         1430 non-null  float64
5   total_sales            1430 non-null  float64
6   tax_status             716 non-null   object
7   tax_class              0 non-null     float64
8   post_author            1430 non-null  float64

```

| | | | |
|----|-----------------------|---------------|---------|
| 9 | post_date | 1430 non-null | object |
| 10 | post_date_gmt | 1430 non-null | object |
| 11 | post_content | 0 non-null | float64 |
| 12 | product_type | 1429 non-null | object |
| 13 | post_title | 1430 non-null | object |
| 14 | post_excerpt | 716 non-null | object |
| 15 | post_status | 1430 non-null | object |
| 16 | comment_status | 1430 non-null | object |
| 17 | ping_status | 1430 non-null | object |
| 18 | post_password | 0 non-null | float64 |
| 19 | post_name | 1430 non-null | object |
| 20 | post_modified | 1430 non-null | object |
| 21 | post_modified_gmt | 1430 non-null | object |
| 22 | post_content_filtered | 0 non-null | float64 |
| 23 | post_parent | 1430 non-null | float64 |
| 24 | guid | 1430 non-null | object |
| 25 | menu_order | 1430 non-null | float64 |
| 26 | post_type | 1430 non-null | object |
| 27 | post_mime_type | 714 non-null | object |
| 28 | comment_count | 1430 non-null | float64 |

dtypes: float64(10), int64(3), object(16)
memory usage: 342.9+ KB

0.7 Vérification des doublons

```
[45]: doublon_web = web.duplicated().sum()
print(f" le table contient {doublon_web} doublon(s) ")
```

le table contient 82 doublon(s)

Anomalie détectée

On constate la présence de 82 doublons dans la table web

0.8 Vérification des valeurs nulles

```
[50]: web.isnull().sum()
```

```
[50]: sku                85
virtual                0
downloadable           0
rating_count           0
average_rating        83
total_sales            83
tax_status            797
tax_class            1513
post_author           83
post_date             83
post_date_gmt         83
```

| | |
|-----------------------|-------|
| post_content | 1513 |
| product_type | 84 |
| post_title | 83 |
| post_excerpt | 797 |
| post_status | 83 |
| comment_status | 83 |
| ping_status | 83 |
| post_password | 1513 |
| post_name | 83 |
| post_modified | 83 |
| post_modified_gmt | 83 |
| post_content_filtered | 1513 |
| post_parent | 83 |
| guid | 83 |
| menu_order | 83 |
| post_type | 83 |
| post_mime_type | 799 |
| comment_count | 83 |
| dtype: | int64 |

Anomalie détectée

Après une vérification des valeurs nulles de la table web le résultat est hilarant, plusieurs valeurs manquantes à l'appel. la colonne sku qui représente en quelque sorte les id des produits sur le web contient 85 valeurs manquantes, des colonnes comme : tax_class, post_content, post_password, post_content_filtered, contiennent 1513 valeurs manquantes sur 1513 lignes. Grosso modo la table web contient plusieurs valeurs manquantes.

```
[54]: total_valeur_manquante = (web.isnull().sum().sum() / web.size ) * 100
      print(f"la table contient {total_valeur_manquante:.2f} % de valeurs manquantes,")
      ↵
```

la table contient 22.85 % de valeurs manquantes

```
[ ]:
```

Une copie de la table s'impose pour un traitement poussé

```
[195]: web_copy = web.copy()
```

```
[197]: LWC, CWC = web_copy.shape
      print(f"la table contient {LWC} ligne(s) et {CWC} colonnes(s) avant suppression,")
      ↵des doublons")
```

la table contient 1513 ligne(s) et 29 colonnes(s) avant suppression des doublons

- Suppression des doublons

```
[200]: web_copy.drop_duplicates(inplace = True)
```

```
[202]: LWC,CWC = web_copy.shape
print(f"la table contient {LWC} ligne(s) et {CWC} colonnes(s) après suppression_
↳des doublons")
```

la table contient 1431 ligne(s) et 29 colonnes(s) après suppression des doublons

- Vérification des valeurs nulles

```
[205]: web_copy.isnull().sum()
```

```
[205]: sku                                3
virtual                                0
downloadable                          0
rating_count                          0
average_rating                        1
total_sales                           1
tax_status                            715
tax_class                             1431
post_author                           1
post_date                             1
post_date_gmt                         1
post_content                          1431
product_type                          2
post_title                            1
post_excerpt                          715
post_status                           1
comment_status                        1
ping_status                           1
post_password                         1431
post_name                             1
post_modified                         1
post_modified_gmt                     1
post_content_filtered                 1431
post_parent                           1
guid                                  1
menu_order                            1
post_type                             1
post_mime_type                        717
comment_count                         1
dtype: int64
```

Après suppression des doublons la table contient encore des valeurs nulles. Des colonnes comme : tax_class , post_content , post_password , post_content_filtered ont la totalité des valeurs manquantes

```
[ ]:
```

Attardons nous un peu sur la colonne sku car c'est cette colonne qu'on veut matcher lors de la lison avec la table de liaison


```
[210]: web_copy[web_copy['sku'].isnull()]
```

```
[210]:
```

| | sku | virtual | downloadable | rating_count | average_rating | total_sales | \ |
|------|-----|---------|--------------|--------------|----------------|-------------|---|
| 8 | NaN | 0 | 0 | 0 | NaN | NaN | |
| 1084 | NaN | 0 | 0 | 0 | 0.0 | -56.0 | |
| 1087 | NaN | 0 | 0 | 0 | 0.0 | -17.0 | |

| | tax_status | tax_class | post_author | post_date | ... | \ |
|------|------------|-----------|-------------|---------------------|-----|---|
| 8 | NaN | NaN | NaN | NaN | ... | |
| 1084 | taxable | NaN | 2.0 | 2018-08-08 11:23:43 | ... | |
| 1087 | taxable | NaN | 2.0 | 2018-07-31 12:07:23 | ... | |

| | post_name | post_modified | \ |
|------|--|---------------------|---|
| 8 | NaN | NaN | |
| 1084 | pierre-jean-villa-condrieu-suspendu-2018 | 2019-11-02 13:24:01 | |
| 1087 | pierre-jean-villa-cote-rotie-fongeant-2017 | 2019-11-02 13:24:15 | |

| | post_modified_gmt | post_content_filtered | post_parent | \ |
|------|---------------------|-----------------------|-------------|---|
| 8 | NaN | NaN | NaN | |
| 1084 | 2019-11-02 12:24:01 | NaN | 0.0 | |
| 1087 | 2019-11-02 12:24:15 | NaN | 0.0 | |

| | guid | menu_order | post_type | \ |
|------|---|------------|-----------|---|
| 8 | NaN | NaN | NaN | |
| 1084 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | |
| 1087 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | |

| | post_mime_type | comment_count |
|------|----------------|---------------|
| 8 | NaN | NaN |
| 1084 | NaN | 0.0 |
| 1087 | NaN | 0.0 |


```
[3 rows x 29 columns]
```

Anomalie détectée

On constate que, dans les trois lignes sans SKU, une ligne est totalement vide, tandis que les deux autres sont presque remplies mais manquent de SKU

```
[214]: web_copy['sku'].map(type).value_counts()
```

```
[214]: sku
<class 'str'>      1428
<class 'float'>    3
Name: count, dtype: int64
```

Anomalie détectée

On constate que la colonne sku n'est pas typée correctement elle contient des valeurs chaîne de

caractère et des valeurs float ce qui n'est pas bon

- Tri par sku

```
[219]: web_copy.sort_values(by='sku')
```

```
[219]:
```

| | sku | virtual | downloadable | rating_count | \ |
|------|---------------------|---------|--------------|--------------|---|
| 668 | 10014 | 0 | 0 | 0 | |
| 1030 | 10014 | 0 | 0 | 0 | |
| 887 | 10459 | 0 | 0 | 0 | |
| 748 | 10459 | 0 | 0 | 0 | |
| 802 | 10775 | 0 | 0 | 0 | |
| ... | ... | ... | ... | ... | |
| 842 | bon-cadeau-25-euros | 0 | 0 | 0 | |
| 1387 | bon-cadeau-25-euros | 0 | 0 | 0 | |
| 8 | NaN | 0 | 0 | 0 | |
| 1084 | NaN | 0 | 0 | 0 | |
| 1087 | NaN | 0 | 0 | 0 | |

| | average_rating | total_sales | tax_status | tax_class | post_author | \ |
|------|----------------|-------------|------------|-----------|-------------|---|
| 668 | 0.0 | 10.0 | taxable | NaN | 2.0 | |
| 1030 | 0.0 | 10.0 | NaN | NaN | 2.0 | |
| 887 | 0.0 | 4.0 | NaN | NaN | 2.0 | |
| 748 | 0.0 | 4.0 | taxable | NaN | 2.0 | |
| 802 | 0.0 | 6.0 | taxable | NaN | 2.0 | |
| ... | ... | ... | ... | ... | ... | |
| 842 | 0.0 | 7.0 | NaN | NaN | 1.0 | |
| 1387 | 0.0 | 7.0 | taxable | NaN | 1.0 | |
| 8 | NaN | NaN | NaN | NaN | NaN | |
| 1084 | 0.0 | -56.0 | taxable | NaN | 2.0 | |
| 1087 | 0.0 | -17.0 | taxable | NaN | 2.0 | |

| | post_date | ... | \ |
|------|---------------------|-----|---|
| 668 | 2019-04-04 15:45:23 | ... | |
| 1030 | 2019-04-04 15:45:23 | ... | |
| 887 | 2018-04-13 15:58:19 | ... | |
| 748 | 2018-04-13 15:58:19 | ... | |
| 802 | 2018-04-17 21:28:52 | ... | |
| ... | ... | ... | |
| 842 | 2018-06-01 13:53:46 | ... | |
| 1387 | 2018-06-01 13:53:46 | ... | |
| 8 | NaN | ... | |
| 1084 | 2018-08-08 11:23:43 | ... | |
| 1087 | 2018-07-31 12:07:23 | ... | |

| | post_name | post_modified | \ |
|------|----------------------------------|---------------------|---|
| 668 | darnleys-london-dry-gin-original | 2020-08-26 15:55:02 | |
| 1030 | darnleys-london-dry-gin-original | 2020-08-26 15:55:02 | |

| | | |
|------|---|---------------------|
| 887 | alphonse-mellot-sancerre-rouge-generation-xix-... | 2020-08-24 14:05:02 |
| 748 | alphonse-mellot-sancerre-rouge-generation-xix-... | 2020-08-24 14:05:02 |
| 802 | albert-mann-pinot-gris-vendanges-tardives-alte... | 2018-12-22 10:30:02 |
| ... | ... | ... |
| 842 | bon-cadeau-de-25-euros | 2018-06-01 14:13:57 |
| 1387 | bon-cadeau-de-25-euros | 2018-06-01 14:13:57 |
| 8 | NaN | NaN |
| 1084 | pierre-jean-villa-condrieu-suspendu-2018 | 2019-11-02 13:24:01 |
| 1087 | pierre-jean-villa-cote-rotie-fongeant-2017 | 2019-11-02 13:24:15 |

| | post_modified_gmt | post_content_filtered | post_parent | \ |
|------|---------------------|-----------------------|-------------|---|
| 668 | 2020-08-26 13:55:02 | NaN | 0.0 | |
| 1030 | 2020-08-26 13:55:02 | NaN | 0.0 | |
| 887 | 2020-08-24 12:05:02 | NaN | 0.0 | |
| 748 | 2020-08-24 12:05:02 | NaN | 0.0 | |
| 802 | 2018-12-22 09:30:02 | NaN | 0.0 | |
| ... | ... | ... | ... | |
| 842 | 2018-06-01 12:13:57 | NaN | 0.0 | |
| 1387 | 2018-06-01 12:13:57 | NaN | 0.0 | |
| 8 | NaN | NaN | NaN | |
| 1084 | 2019-11-02 12:24:01 | NaN | 0.0 | |
| 1087 | 2019-11-02 12:24:15 | NaN | 0.0 | |

| | guid | menu_order | \ |
|------|---|------------|---|
| 668 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 1030 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| 887 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| 748 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 802 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| ... | ... | ... | |
| 842 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| 1387 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 8 | NaN | NaN | |
| 1084 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 1087 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |

| | post_type | post_mime_type | comment_count |
|------|------------|----------------|---------------|
| 668 | product | NaN | 0.0 |
| 1030 | attachment | image/jpeg | 0.0 |
| 887 | attachment | image/jpeg | 0.0 |
| 748 | product | NaN | 0.0 |
| 802 | product | NaN | 0.0 |
| ... | ... | ... | ... |
| 842 | attachment | image/jpeg | 0.0 |
| 1387 | product | NaN | 0.0 |
| 8 | NaN | NaN | NaN |
| 1084 | product | NaN | 0.0 |

```
1087      product      NaN      0.0
```

```
[1431 rows x 29 columns]
```

Anomalie détectée

On constate que des valeurs on des sku **‘bon-cadeau-25-euros’** ce qui confirme un réel problème de typage dans la colonne sku

```
[ ]:
```

- Uniformisation des valeurs

```
[225]: web_copy['sku'] = web_copy['sku'].astype(str)
```

On uniformise tout pour avoir un seul type L'étape `astype(str)` garantit que toutes les entrées de la colonne sont traitées comme des chaînes de caractères, éliminant toute ambiguïté ou problème qui pourrait survenir si certaines valeurs étaient traitées comme des nombres et d'autres comme du texte.

```
[228]: web_copy.isnull().sum()
```

```
[228]: sku                0
virtual                0
downloadable          0
rating_count          0
average_rating        1
total_sales           1
tax_status            715
tax_class            1431
post_author           1
post_date             1
post_date_gmt         1
post_content         1431
product_type          2
post_title            1
post_excerpt          715
post_status           1
comment_status        1
ping_status           1
post_password         1431
post_name             1
post_modified         1
post_modified_gmt     1
post_content_filtered 1431
post_parent           1
guid                 1
menu_order            1
post_type             1
```

```
post_mime_type      717
comment_count        1
dtype: int64
```

```
[ ]:
```

- Suppression des colonnes complètement vides

```
[232]: colonne_s = ['tax_class' , 'post_content' , 'post_password' ,
↳ 'post_content_filtered' ]

web_copy.drop( columns = colonne_s , inplace = True)
```

```
[234]: web_copy.isnull().sum()
```

```
[234]: sku      0
virtual      0
downloadable  0
rating_count  0
average_rating  1
total_sales   1
tax_status    715
post_author   1
post_date     1
post_date_gmt  1
product_type   2
post_title    1
post_excerpt  715
post_status    1
comment_status 1
ping_status    1
post_name      1
post_modified  1
post_modified_gmt 1
post_parent    1
guid           1
menu_order     1
post_type      1
post_mime_type 717
comment_count   1
dtype: int64
```

```
[ ]:
```

```
[237]: LISSTE = web_copy.sort_values(by='sku',ascending = False)
LISSTE
```

[237]:

| | sku | virtual | downloadable | rating_count | \ |
|------|---------------------|---------|--------------|--------------|---|
| 1084 | nan | 0 | 0 | 0 | |
| 8 | nan | 0 | 0 | 0 | |
| 1087 | nan | 0 | 0 | 0 | |
| 1387 | bon-cadeau-25-euros | 0 | 0 | 0 | |
| 842 | bon-cadeau-25-euros | 0 | 0 | 0 | |
| ... | ... | ... | ... | ... | |
| 1317 | 10775 | 0 | 0 | 0 | |
| 748 | 10459 | 0 | 0 | 0 | |
| 887 | 10459 | 0 | 0 | 0 | |
| 668 | 10014 | 0 | 0 | 0 | |
| 1030 | 10014 | 0 | 0 | 0 | |

| | average_rating | total_sales | tax_status | post_author | \ |
|------|----------------|-------------|------------|-------------|---|
| 1084 | 0.0 | -56.0 | taxable | 2.0 | |
| 8 | NaN | NaN | NaN | NaN | |
| 1087 | 0.0 | -17.0 | taxable | 2.0 | |
| 1387 | 0.0 | 7.0 | taxable | 1.0 | |
| 842 | 0.0 | 7.0 | NaN | 1.0 | |
| ... | ... | ... | ... | ... | |
| 1317 | 0.0 | 6.0 | NaN | 2.0 | |
| 748 | 0.0 | 4.0 | taxable | 2.0 | |
| 887 | 0.0 | 4.0 | NaN | 2.0 | |
| 668 | 0.0 | 10.0 | taxable | 2.0 | |
| 1030 | 0.0 | 10.0 | NaN | 2.0 | |

| | post_date | post_date_gmt | ... | ping_status | \ |
|------|---------------------|---------------------|-----|-------------|---|
| 1084 | 2018-08-08 11:23:43 | 2018-08-08 09:23:43 | ... | closed | |
| 8 | NaN | NaN | ... | NaN | |
| 1087 | 2018-07-31 12:07:23 | 2018-07-31 10:07:23 | ... | closed | |
| 1387 | 2018-06-01 13:53:46 | 2018-06-01 11:53:46 | ... | closed | |
| 842 | 2018-06-01 13:53:46 | 2018-06-01 11:53:46 | ... | closed | |
| ... | ... | ... | ... | ... | |
| 1317 | 2018-04-17 21:28:52 | 2018-04-17 19:28:52 | ... | closed | |
| 748 | 2018-04-13 15:58:19 | 2018-04-13 13:58:19 | ... | closed | |
| 887 | 2018-04-13 15:58:19 | 2018-04-13 13:58:19 | ... | closed | |
| 668 | 2019-04-04 15:45:23 | 2019-04-04 13:45:23 | ... | closed | |
| 1030 | 2019-04-04 15:45:23 | 2019-04-04 13:45:23 | ... | closed | |

| | post_name | post_modified | \ |
|------|---|---------------------|---|
| 1084 | pierre-jean-villa-condrieu-suspendu-2018 | 2019-11-02 13:24:01 | |
| 8 | NaN | NaN | |
| 1087 | pierre-jean-villa-cote-rotie-fongeant-2017 | 2019-11-02 13:24:15 | |
| 1387 | bon-cadeau-de-25-euros | 2018-06-01 14:13:57 | |
| 842 | bon-cadeau-de-25-euros | 2018-06-01 14:13:57 | |
| ... | ... | ... | |
| 1317 | albert-mann-pinot-gris-vendanges-tardives-alte... | 2018-12-22 10:30:02 | |

| | | |
|------|---|---------------------|
| 748 | alphonse-mellot-sancerre-rouge-generation-xix-... | 2020-08-24 14:05:02 |
| 887 | alphonse-mellot-sancerre-rouge-generation-xix-... | 2020-08-24 14:05:02 |
| 668 | darnleys-london-dry-gin-original | 2020-08-26 15:55:02 |
| 1030 | darnleys-london-dry-gin-original | 2020-08-26 15:55:02 |

| | post_modified_gmt | post_parent | \ |
|------|---------------------|-------------|---|
| 1084 | 2019-11-02 12:24:01 | 0.0 | |
| 8 | NaN | NaN | |
| 1087 | 2019-11-02 12:24:15 | 0.0 | |
| 1387 | 2018-06-01 12:13:57 | 0.0 | |
| 842 | 2018-06-01 12:13:57 | 0.0 | |
| ... | ... | ... | |
| 1317 | 2018-12-22 09:30:02 | 0.0 | |
| 748 | 2020-08-24 12:05:02 | 0.0 | |
| 887 | 2020-08-24 12:05:02 | 0.0 | |
| 668 | 2020-08-26 13:55:02 | 0.0 | |
| 1030 | 2020-08-26 13:55:02 | 0.0 | |

| | guid | menu_order | \ |
|------|---|------------|---|
| 1084 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 8 | NaN | NaN | |
| 1087 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 1387 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 842 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| ... | ... | ... | |
| 1317 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| 748 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 887 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |
| 668 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | |
| 1030 | https://www.bottle-neck.fr/wp-content/uploads/... | 0.0 | |

| | post_type | post_mime_type | comment_count |
|------|------------|----------------|---------------|
| 1084 | product | NaN | 0.0 |
| 8 | NaN | NaN | NaN |
| 1087 | product | NaN | 0.0 |
| 1387 | product | NaN | 0.0 |
| 842 | attachment | image/jpeg | 0.0 |
| ... | ... | ... | ... |
| 1317 | attachment | image/jpeg | 0.0 |
| 748 | product | NaN | 0.0 |
| 887 | attachment | image/jpeg | 0.0 |
| 668 | product | NaN | 0.0 |
| 1030 | attachment | image/jpeg | 0.0 |

[1431 rows x 25 columns]

Anomalie détectée

Il semblerait aussi que chaque élément apparaîtrait deux fois , une première fois en tant que produit et une deuxième fois avec une image sûrement pour le site web. Cela peut se confirmer via la colonne `post_type` qui a deux valeurs différentes pour un même produit , une avec `product` et l'autre avec `attachment` suivi de la colonne `post_mime_type` ou est affiché `image/jpeg` .

```
[ ]:
```

Analyse de la table `LIAISON`.

```
[243]: ligne_L , colonne_L = liaison.shape

print(f" la table liaison contient {ligne_L} ligne(s) et {colonne_L} colonne(s) ")
↪
```

la table `liaison` contient 825 ligne(s) et 2 colonne(s)

```
[245]: liaison.isnull().sum()
```

```
[245]: id_web      91
product_id     0
dtype: int64
```

Anomalie détectée

Après vérification, On constate la présence de valeurs manquantes et la majorité sinon l'unique absence provient de la colonne des `id web` qui fait référence aux données provenant du web . La colonne `product_id` qui fait référence aux données provenant de ERP ne contient pas de valeurs manquantes .

```
[249]: V_manquante_L = ( liaison.isnull().sum().sum() / liaison.size ) * 100
print(f" La proportion de valeur manquante dans la table liaison est de ")
↪ {V_manquante_L:.1f} % )
```

La proportion de valeur manquante dans la table `liaison` est de 5.5 %

```
[ ]:
```

```
[252]: liaison.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 825 entries, 0 to 824
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   id_web      734 non-null   object
1   product_id  825 non-null   int64
dtypes: int64(1), object(1)
memory usage: 13.0+ KB
```

```
[ ]:
```


- Un constat est fait , la table ERP contient 825 lignes comme la table liaison, une eventuel jointure peut etre faite entre ERP et LIAISON surtout que l'absence de valeur dans la table liaison ne provient pas des product_id qui font référence à la table ERP

Stratégie de jointure des données.

0.8.1 Jointure entre ERP & LIAISON

```
[258]: erp_liaison = erp.merge( liaison , on ='product_id' )
```

```
[260]: erp_liaison
```

```
[260]:
```

| | product_id | onsale_web | price | stock_quantity | stock_status | \ |
|-----|------------|------------|-------|----------------|--------------|---|
| 0 | 3847 | 1 | 24.2 | 16 | instock | |
| 1 | 3849 | 1 | 34.3 | 10 | instock | |
| 2 | 3850 | 1 | 20.8 | 0 | outofstock | |
| 3 | 4032 | 1 | 14.1 | 26 | instock | |
| 4 | 4039 | 1 | 46.0 | 3 | outofstock | |
| .. | ... | ... | ... | ... | ... | |
| 820 | 7203 | 0 | 45.0 | 30 | instock | |
| 821 | 7204 | 0 | 45.0 | 9 | instock | |
| 822 | 7247 | 1 | 54.8 | 6 | instock | |
| 823 | 7329 | 0 | 26.5 | 14 | instock | |
| 824 | 7338 | 1 | 16.3 | 40 | instock | |

| | purchase_price | id_web |
|-----|----------------|---------|
| 0 | 12.88 | 15298 |
| 1 | 17.54 | 15296 |
| 2 | 10.64 | 15300 |
| 3 | 6.92 | 19814 |
| 4 | 23.77 | 19815 |
| .. | ... | ... |
| 820 | 23.48 | NaN |
| 821 | 24.18 | NaN |
| 822 | 27.18 | 13127-1 |
| 823 | 13.42 | 14680-1 |
| 824 | 8.00 | 16230 |

[825 rows x 7 columns]

0.9 Vérification des valeurs nulles

```
[263]: erp_liaison.isnull().sum()
```

```
[263]: product_id      0
      onsale_web      0
      price          0
      stock_quantity  0
```

```
stock_status      0
purchase_price    0
id_web            91
dtype: int64
```

[]:

0.10 Jointure entre `erp_liaison` & `WEB`

```
[267]: erp_liaison_web = erp_liaison.dropna().merge(web.loc[web['post_type'] == 'product'], left_on='id_web', right_on='sku' )
```

- **`erp_liaison.dropna()`** : La méthode `.dropna()` est appliquée pour supprimer toutes les lignes de ce DataFrame qui contiennent au moins une valeur manquante (NaN). L'objectif est de garantir que seules les informations complètes et valides de l'ERP sont utilisées pour la suite de l'opération, ce qui est crucial pour la qualité de la jointure.
- **`web.loc[web['post_type'] == 'product']`** : Cette expression sélectionne uniquement les lignes du DataFrame `web` où la colonne `post_type` a spécifiquement la valeur 'product'. Ceci est essentiel pour s'assurer que nous fusionnons uniquement des entrées qui correspondent à des produits réels et vendables de notre site web, en excluant d'autres types de contenu (comme 'attachment').
- **`.merge(... , left_on='id_web', right_on='sku')`** : `left_on='id_web'` : Indique que la colonne `id_web` du DataFrame de gauche (le `erp_liaison` nettoyé) sera utilisée comme clé principale pour la jointure.
- **`right_on='sku'`** : Spécifie que la colonne `sku` du DataFrame de droite (le `web` filtré) sera la clé correspondante. Le `sku` (Stock Keeping Unit) est l'identifiant unique pour les produits dans la table `web`.

Cette méthode effectue une inner join. Cela signifie que le nouveau DataFrame `erp_liaison_web` sera créé en incluant uniquement les lignes pour lesquelles une correspondance exacte a été trouvée entre un `id_web` de l'ERP et un `sku` d'un produit du site web.

0.11 Vérification des dimensions

```
[271]: l, c = erp_liaison_web.shape
print(f" le DataFrame contient {l} ligne(s) et {c} colonne(s) ")
```

le DataFrame contient 714 ligne(s) et 36 colonne(s)

0.12 Vérification des valeurs nulles

```
[274]: erp_liaison_web.isnull().sum()
```

```
[274]: product_id      0
onsale_web           0
price                0
stock_quantity       0
```

```

stock_status      0
purchase_price    0
id_web            0
sku              0
virtual          0
downloadable      0
rating_count      0
average_rating    0
total_sales       0
tax_status        0
tax_class         714
post_author       0
post_date        0
post_date_gmt    0
post_content      714
product_type      1
post_title        0
post_excerpt      0
post_status       0
comment_status    0
ping_status       0
post_password     714
post_name         0
post_modified     0
post_modified_gmt 0
post_content_filtered 714
post_parent       0
guid             0
menu_order        0
post_type         0
post_mime_type    714
comment_count     0
dtype: int64

```

Sur 714 ligne ces colonnes ont la totalité des valeurs manquantes : `tax_class` , `post_content` , `post_password` , `post_content_filtered` et `post_mime_type`

la colonne `product_type` possède une seule valeur manquante , on peut procéder à une imputation par le mode , pourquoi par le mode , cette colonne est une colonne qualitative .

0.13 Suppression des colonnes totalement vides

```

[279]: colonnes_vides = ['tax_class', 'post_content' , 'post_password' ,
↳ 'post_content_filtered', 'post_mime_type']
erp_liaison_web.drop(columns=colonnes_vides , inplace = True )

```

```

[281]: l,c = erp_liaison_web.shape

```

```
print(f" le DataFrame contient {l} ligne(s) et {c} colonne(s) après suppression_
↳des colonnes vides")
```

le DataFrame contient 714 ligne(s) et 31 colonne(s) après suppression des colonnes vides

```
[283]: erp_liaison_web.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 714 entries, 0 to 713
Data columns (total 31 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            714 non-null   int64
1   onsale_web            714 non-null   int64
2   price                 714 non-null   float64
3   stock_quantity       714 non-null   int64
4   stock_status         714 non-null   object
5   purchase_price       714 non-null   float64
6   id_web               714 non-null   object
7   sku                  714 non-null   object
8   virtual              714 non-null   int64
9   downloadable         714 non-null   int64
10  rating_count         714 non-null   int64
11  average_rating       714 non-null   float64
12  total_sales          714 non-null   float64
13  tax_status           714 non-null   object
14  post_author          714 non-null   float64
15  post_date            714 non-null   object
16  post_date_gmt        714 non-null   object
17  product_type         713 non-null   object
18  post_title           714 non-null   object
19  post_excerpt         714 non-null   object
20  post_status          714 non-null   object
21  comment_status       714 non-null   object
22  ping_status          714 non-null   object
23  post_name            714 non-null   object
24  post_modified        714 non-null   object
25  post_modified_gmt    714 non-null   object
26  post_parent          714 non-null   float64
27  guid                714 non-null   object
28  menu_order           714 non-null   float64
29  post_type            714 non-null   object
30  comment_count        714 non-null   float64
dtypes: float64(8), int64(6), object(17)
memory usage: 173.1+ KB
```

```
[285]: erp_liaison_web['product_type'].unique()
```

```
[285]: array(['Vin', 'Champagne', 'Whisky', 'Cognac', nan, "Huile d'olive",  
          'Gin'], dtype=object)
```

```
[287]: erp_liaison_web['product_type'].value_counts()
```

```
[287]: product_type  
Vin          658  
Champagne    28  
Whisky       14  
Cognac        8  
Huile d'olive 3  
Gin           2  
Name: count, dtype: int64
```

```
[289]: erp_liaison_web['product_type'].describe()
```

```
[289]: count      713  
unique        6  
top      Vin  
freq      658  
Name: product_type, dtype: object
```

- Récupère le mode de la colonne product_type ou il manque qu'une seule valeur, on récupère le mode aussi car cette colonne contient des valeurs qualitatives et il est plus ingénieux de faire une imputation par le mode

```
[292]: mode_product_type = erp_liaison_web['product_type'].mode()[0]  
mode_product_type
```

```
[292]: 'Vin'
```

- Remplace la valeur nulle par le mode

```
[ ]: erp_liaison_web['product_type'].fillna(mode_product_type, inplace = True)
```

```
[297]: erp_liaison_web['product_type'].unique()
```

```
[297]: array(['Vin', 'Champagne', 'Whisky', 'Cognac', "Huile d'olive", 'Gin'],  
          dtype=object)
```

```
[299]: erp_liaison_web.isnull().sum()
```

```
[299]: product_id      0  
onsale_web        0  
price             0  
stock_quantity    0  
stock_status      0  
purchase_price    0  
id_web            0
```

```

sku                0
virtual            0
downloadable       0
rating_count       0
average_rating     0
total_sales        0
tax_status         0
post_author        0
post_date          0
post_date_gmt      0
product_type       0
post_title         0
post_excerpt       0
post_status        0
comment_status     0
ping_status        0
post_name          0
post_modified      0
post_modified_gmt  0
post_parent        0
guid              0
menu_order         0
post_type          0
comment_count      0
dtype: int64

```

Suite à ces manœuvres on peut constater que le dataFrame `erp_liaison_web` est propre

[]:

0.14 Avant d'exploiter les données, un travail de nettoyage a été effectué. Voici les principales anomalies identifiées lors de la phase de consolidation :

| Fichier | Colonne(s) concernée(s) | Type d'erreur | Détail / Explication |
|---------|-------------------------|--------------------|---|
| web | Toutes colonnes | Valeurs manquantes | Plusieurs colonnes (tax_class, post_password, post_content_filtered) contiennent 1513 valeurs manquantes sur 1513 lignes. |

| Fichier | Colonne(s) concernée(s) | Type d'erreur | Détail / Explication |
|---------|--------------------------------|-------------------------|--|
| web | sku | Valeurs manquantes | 85 sku manquants, alors qu'il s'agit d'un identifiant clé du produit. |
| web | Ligne entière | Ligne totalement vide | Une ligne ne contient aucune donnée (vidée ou corrompue). |
| web | sku | Problème de typage | Mélange de float et de string , avec des sku comme 'bon-cadeau-25-euros'. |
| web | sku, post_type, post_mime_type | Doublons logiques | Chaque produit apparaît deux fois : une fois comme product , une fois comme attachment (image du produit). |
| web | — | Doublons réels | 82 doublons détectés avec <code>.duplicated().sum()</code> . |
| liaison | id_web | Valeurs manquantes | Certains produits ERP ne sont pas liés à un id_web , donc introuvables côté web. |
| liaison | product_id | Aucune valeur manquante | Toutes les références ERP sont bien présentes. |

Ces anomalies ont été prises en compte dans le nettoyage des données avant fusion et analyse.

[]:

[]:

Phase 2 : Analyse des données et reporting

```
[307]: aL, bC = erp_liaison_web.shape
print(f" le dataframe nettoy  comprend {aL} ligne(s) et {bC} colonne(s) ")
```

le dataframe nettoy  comprend 714 ligne(s) et 31 colonne(s)

```
[309]: erp_liaison_web.columns
```

```
[309]: Index(['product_id', 'onsale_web', 'price', 'stock_quantity', 'stock_status',
        'purchase_price', 'id_web', 'sku', 'virtual', 'downloadable',
        'rating_count', 'average_rating', 'total_sales', 'tax_status',
        'post_author', 'post_date', 'post_date_gmt', 'product_type',
        'post_title', 'post_excerpt', 'post_status', 'comment_status',
        'ping_status', 'post_name', 'post_modified', 'post_modified_gmt',
        'post_parent', 'guid', 'menu_order', 'post_type', 'comment_count'],
        dtype='object')
```

```
[ ]:
```

0.15 Calculons le chiffre d'affaire par produit

```
[313]: erp_liaison_web['CA'] = erp_liaison_web['price'] *
        ↪erp_liaison_web['total_sales']
tableau_ca =
        ↪erp_liaison_web[['product_id', 'post_name', 'price', 'purchase_price', 'total_sales',
        ↪, 'CA']]

print("Les chiffre d'affaire par produit : ")
tableau_ca.head(10)
```

Les chiffre d'affaire par produit :

```
[313]:
```

| | product_id | post_name | price | \ |
|---|------------|---|-------|---|
| 0 | 3847 | pierre-jean-villa-saint-joseph-preface-2018 | 24.2 | |
| 1 | 3849 | pierre-jean-villa-saint-joseph-tilde-2017 | 34.3 | |
| 2 | 3850 | pierre-jean-villa-croze-hermitage-accroche-coe... | 20.8 | |
| 3 | 4032 | pierre-jean-villa-igp-gamine-2018 | 14.1 | |
| 4 | 4039 | pierre-jean-villa-cote-rotie-carmina-2017 | 46.0 | |
| 5 | 4040 | pierre-jean-villa-saint-joseph-saut-ange-2018 | 34.3 | |
| 6 | 4041 | pierre-gaillard-condrieu-2018 | 32.7 | |
| 7 | 4042 | pierre-gaillard-cornas-2017 | 31.2 | |
| 8 | 4043 | pierre-gaillard-cote-rotie-esprit-blond-2017 | 60.0 | |
| 9 | 4045 | pierre-gaillard-cote-rotie-2018 | 42.6 | |

| | purchase_price | total_sales | CA |
|---|----------------|-------------|-------|
| 0 | 12.88 | 6.0 | 145.2 |
| 1 | 17.54 | 9.0 | 308.7 |
| 2 | 10.64 | 0.0 | 0.0 |
| 3 | 6.92 | 12.0 | 169.2 |

| | | | |
|---|-------|-----|-------|
| 4 | 23.77 | 3.0 | 138.0 |
| 5 | 18.25 | 7.0 | 240.1 |
| 6 | 17.57 | 6.0 | 196.2 |
| 7 | 15.48 | 8.0 | 249.6 |
| 8 | 29.45 | 0.0 | 0.0 |
| 9 | 22.01 | 3.0 | 127.8 |

0.16 Calculons le chiffre d'affaires total général.

```
[316]: CA_TOTAL = erp_liaison_web['CA'].sum()
print(f" Le chiffre d'affaire Total : {CA_TOTAL:,.2f} FCFA".format().
      ↪replace(',', ' '))
```

Le chiffre d'affaire Total : 143 680.10 FCFA

[]:

0.17 Identifier les produits les plus performants, appliquer la règle du 20/80

- On effectue d'abord un tri pour voir les produits les plus rentables

```
[321]: Ensemble_des_produits = erp_liaison_web.sort_values(by='CA', ascending = False)
```

- On calcul combien chaque produit rapporte au CA , ensuite on cumule ces apports pour voir la zone des 80 %

```
[324]: Ensemble_des_produits['part %'] = Ensemble_des_produits['CA'] / CA_TOTAL
Ensemble_des_produits['Cumule %'] = Ensemble_des_produits['part %'].cumsum()
```

- On applique la loi de pareto qui stipule que 80% du chiffre d'affaire est produit 20% des produits. On identifie donc les produits qui cumulent 80% du CA et on les marque à l'aide d'une colonne booléenne

```
[327]: Ensemble_des_produits['Top 20% Produit'] = Ensemble_des_produits['Cumule %'] <=
      ↪0.80
```

```
[329]: TOTAL_PRODUIT_COUNT = Ensemble_des_produits['Top 20% Produit'].sum()

TOTAL_PRODUIT_DF = len(Ensemble_des_produits)

POURCENTAGE_80_P = TOTAL_PRODUIT_COUNT / TOTAL_PRODUIT_DF

print("Pourcentage de produits représentant 80% du CA :",
      ↪round(POURCENTAGE_80_P * 100, 2), "%")
```

Pourcentage de produits représentant 80% du CA : 60.78 %

[]:

- Le nombre de produit présent dans le DATAFRAME

```
[333]: print(f" le nombre de produit présent:  {TOTAL_PRODUIT_DF}")
```

le nombre de produit présent: 714

```
[335]: print(f" le nombre de produit qui contribue à hauteur de 80% au CA:␣
↪{TOTAL_PRODUIT_COUNT}")
```

le nombre de produit qui contribue à hauteur de 80% au CA: 434

```
[494]: sizes_ = [TOTAL_PRODUIT_COUNT , TOTAL_PRODUIT_DF - TOTAL_PRODUIT_COUNT ]
names_ = ['Top 80% CA', 'Autres Produits']

fig = px.pie(values = sizes_ ,
              names = names_ ,
              title='Répartition des produits : Top 80% du CA vs Autres',
              color_discrete_sequence=px.colors.sequential.RdBu)
fig.show()
```

Répartition des produits : Top 80% du CA vs Autres



0.18 Le résultat montre que 80 % du chiffre d’affaire est généré par environ 61 % de produit .Ainsi, la loi de pareto n’est pas strictement respectée ce qui peut indiquer une offre de prix des produits relativement homogene

- AFFICHAGE DE NOS PRODUITS PRODUISANT 80% DU CHIFFRE D’AFFAIRE

```
[340]: Ensemble_des_produits.loc[
        Ensemble_des_produits['Top 20% Produit'] == True,
        ['product_id', 'post_name', 'price', 'purchase_price', 'total_sales', 'CA'],␣
        ↪'part %', 'Cumule %', 'Top 20% Produit']
        ].sort_values(by='Cumule %',ascending = False)
```

```
[340]:   product_id   post_name  price  \
364      4865  mourgues-du-gres-costieres-de-nimes-capitelles...    9.8
37       4075      plateau-chenes-lirac-2015    14.7
323      4728      schoech-gewurztraminer-vt-2017    29.5
277      4664  la-cotelleraie-saint-nicolas-de-bourgueil-le-v...    16.4
```

| | | | |
|-----|------|---|-------|
| 420 | 4978 | chateau-dutruch-grand-poujeaux-moulis-2016 | 18.5 |
| .. | ... | ... | ... |
| 653 | 6212 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 115.0 |
| 582 | 5826 | agnes-levet-amethyste-2017 | 41.2 |
| 200 | 4353 | champagne-egly-ouriet-grand-cru-brut-rose | 79.5 |
| 587 | 5892 | coteaux-champenois-egly-ouriet-ambonnay-rouge-... | 191.3 |
| 199 | 4352 | champagne-egly-ouriet-grand-cru-millesime-2008 | 225.0 |

| | purchase_price | total_sales | CA | part % | Cumule % | Top 20% Produit |
|-----|----------------|-------------|--------|----------|----------|-----------------|
| 364 | 5.16 | 15.0 | 147.0 | 0.001023 | 0.799839 | True |
| 37 | 7.97 | 10.0 | 147.0 | 0.001023 | 0.798816 | True |
| 323 | 14.48 | 5.0 | 147.5 | 0.001027 | 0.797792 | True |
| 277 | 8.39 | 9.0 | 147.6 | 0.001027 | 0.796766 | True |
| 420 | 9.46 | 8.0 | 148.0 | 0.001030 | 0.795739 | True |
| .. | ... | ... | ... | ... | ... | ... |
| 653 | 59.42 | 7.0 | 805.0 | 0.005603 | 0.044298 | True |
| 582 | 21.71 | 20.0 | 824.0 | 0.005735 | 0.038696 | True |
| 200 | 45.91 | 14.0 | 1113.0 | 0.007746 | 0.032961 | True |
| 587 | 116.06 | 6.0 | 1147.8 | 0.007989 | 0.025214 | True |
| 199 | 137.81 | 11.0 | 2475.0 | 0.017226 | 0.017226 | True |

[434 rows x 9 columns]

- Comparaison entre ceux qui rapportent en terme de CA et ceux qui rapportent en terme de vente

```
[342]: top_10_CA = Ensemble_des_produits.loc[Ensemble_des_produits['Top 20% Produit']
↳ == True].sort_values(by='CA', ascending=False).head(10)

fig = px.bar( top_10_CA, y='post_name' , x='CA' ,
               text= top_10_CA['CA'].round(1),
               title='Les 10 produits avec de gros CA unique',
               labels= {'post_name': 'Produit', 'CA': 'Chiffre d'affaires'} )

fig.update_layout(
    width=1000,
    height=600,
)
fig.show()
```

```
[343]: top_10_Vente = Ensemble_des_produits.sort_values(by='total_sales',
↳ ascending=False).head(10)

fig = px.bar( top_10_Vente , y='post_name' , x='total_sales' ,
               text='total_sales',
               title='Les 10 produits avec les grands nombre de vente',
```

```

        labels= {'post_name':'Produit','total_sales':'Nombre de vente'} )

fig.update_layout(
    width=1000,
    height=600,
)
fig.show()

```

[]:

```

[345]: fig = make_subplots(
        rows=2, cols=1,
        subplot_titles=(
            "Les 10 produits avec de gros CA unique",
            "Les 10 produits avec les grands nombre de vente"
        )
    )

fig.add_trace(
    go.Bar(
        y=top_10_CA['post_name'],
        x=top_10_CA['CA'],
        orientation='h',
        name='CA',
        marker_color='steelblue',
        text=top_10_CA['CA'].round(1),
        textposition='auto'
    ),
    row=1, col=1
)

fig.add_trace(
    go.Bar(
        y=top_10_Vente['post_name'],
        x=top_10_Vente['total_sales'],
        orientation='h',
        name='Ventes',
        marker_color='darkorange',
        text=top_10_Vente['total_sales'],
        textposition='auto'
    ),
    row=2, col=1
)

fig.update_layout(

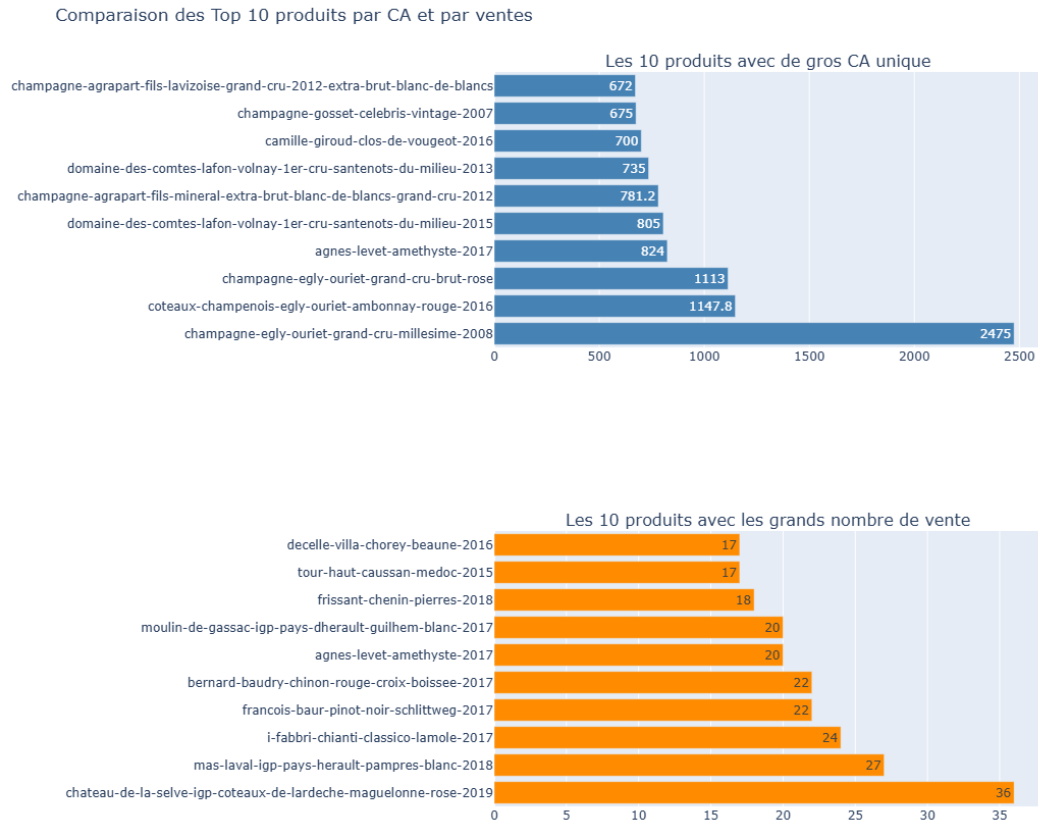
```

```

width=1000,
height=900,
title_text='Comparaison des Top 10 produits par CA et par ventes',
showlegend=False,
font=dict(size=12)
)

fig.show()

```



[]:

0.18.1 Dans les produits les plus rentables en termes de vente et de chiffre d'affaires, on va essayer de ressortir ceux qui se démarquent des deux côtés.

```

[353]: print(top_10_Vente[['post_name', 'total_sales']])

print(" ===== " )

```

```
print(top_10_CA[['post_name', 'CA']])
```

| | post_name | total_sales |
|-----|---|-------------|
| 365 | chateau-de-la-selve-igp-coteaux-de-lardeche-ma... | 36.0 |
| 122 | mas-laval-igp-pays-herault-pampres-blanc-2018 | 27.0 |
| 175 | i-fabbri-chianti-classico-lamole-2017 | 24.0 |
| 321 | francois-baur-pinot-noir-schlittweg-2017 | 22.0 |
| 265 | bernard-baudry-chinon-rouge-croix-boissee-2017 | 22.0 |
| 582 | agnes-levet-amethyste-2017 | 20.0 |
| 645 | moulin-de-gassac-igp-pays-dherault-guilhem-bla... | 20.0 |
| 135 | frissant-chenin-pierres-2018 | 18.0 |
| 570 | tour-haut-caussan-medoc-2015 | 17.0 |
| 673 | decelle-villa-chorey-beaune-2016 | 17.0 |

```
=====
```

| | post_name | CA |
|-----|---|--------|
| 199 | champagne-egly-ouriet-grand-cru-millesime-2008 | 2475.0 |
| 587 | coteaux-champenois-egly-ouriet-ambonnay-rouge-... | 1147.8 |
| 200 | champagne-egly-ouriet-grand-cru-brut-rose | 1113.0 |
| 582 | agnes-levet-amethyste-2017 | 824.0 |
| 653 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 805.0 |
| 438 | champagne-agrapart-fils-mineral-extra-brut-bla... | 781.2 |
| 432 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 735.0 |
| 553 | camille-giroud-clos-de-vougeot-2016 | 700.0 |
| 642 | champagne-gosset-celebris-vintage-2007 | 675.0 |
| 437 | champagne-agrapart-fils-lavizoise-grand-cru-20... | 672.0 |

```
[354]: RV = set(top_10_Vente['post_name'])
RC = set(top_10_CA['post_name'])

union_produit_nbr_vente_ca = RV & RC
```

```
[355]: Ensemble_des_produits[Ensemble_des_produits['post_name'].
↳isin(union_produit_nbr_vente_ca)]
```

```
[355]:
```

| | product_id | onsale_web | price | stock_quantity | stock_status | \ |
|-----|------------|------------|-------|----------------|--------------|---|
| 582 | 5826 | 1 | 41.2 | 34 | instock | |

| | purchase_price | id_web | sku | virtual | downloadable | ... | \ |
|-----|----------------|--------|-------|---------|--------------|-----|---|
| 582 | 21.71 | 15325 | 15325 | 0 | 0 | ... | |

| | post_modified_gmt | post_parent | \ |
|-----|---------------------|-------------|---|
| 582 | 2020-05-21 12:00:02 | 0.0 | |

| | guid | menu_order | post_type | \ |
|-----|---|------------|-----------|---|
| 582 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | |

| | comment_count | CA | part % | Cumule % | Top 20% Produit |
|-----|---------------|-------|----------|----------|-----------------|
| 582 | 0.0 | 824.0 | 0.005735 | 0.038696 | True |

```
[1 rows x 35 columns]
```

```
[359]: PP = Ensemble_des_produits[Ensemble_des_produits['product_id'] == 5826 ]
PP[['product_id', 'post_name', 'price', 'purchase_price', 'total_sales', 'CA', 'part %', 'Cumule %', 'Top 20% Produit']]
```

```
[359]:
```

| | product_id | | post_name | price | purchase_price | \ |
|-----|------------|--|----------------------------|-------|----------------|---|
| 582 | 5826 | | agnes-levet-amethyste-2017 | 41.2 | 21.71 | |

| | total_sales | CA | part % | Cumule % | Top 20% Produit |
|-----|-------------|-------|----------|----------|-----------------|
| 582 | 20.0 | 824.0 | 0.005735 | 0.038696 | True |

On observe que certains produits comme le agnes-levet-amethyste-2017 combinent un bon CA et de bonnes ventes. Ce sont des produits à fort levier économique

```
[ ]:
```

0.19 Détection des valeurs aberrantes sur les prix

Suite à l'analyse de la répartition du chiffre d'affaires par produit, nous avons constaté que **la règle de Pareto (80/20)** n'était **pas strictement respectée** dans notre base : environ **61 % des produits** génèrent **80 % du chiffre d'affaires**. Cela peut refléter une certaine homogénéité dans l'offre commerciale ou une stratégie de tarification étalée.

Afin de vérifier si cette distribution est également affectée par **des valeurs aberrantes**, nous avons appliqué une détection d'**outliers** sur la variable **price**, à l'aide des méthodes Z-score et de l'écart interquartile (IQR).

0.19.1 Utilisation du Z-score

```
[365]: erp_liaison_web['z_score'] = zscore(erp_liaison_web['price'])

outliers_des_prix = erp_liaison_web[ (erp_liaison_web['z_score'] > 3) |
    (erp_liaison_web['z_score'] < -3) ]

print(f" En utilisant le z-score on trouve {len(outliers_des_prix)} outliers ")
```

```
En utilisant le z-score on trouve 13 outliers
```

```
[ ]:
```

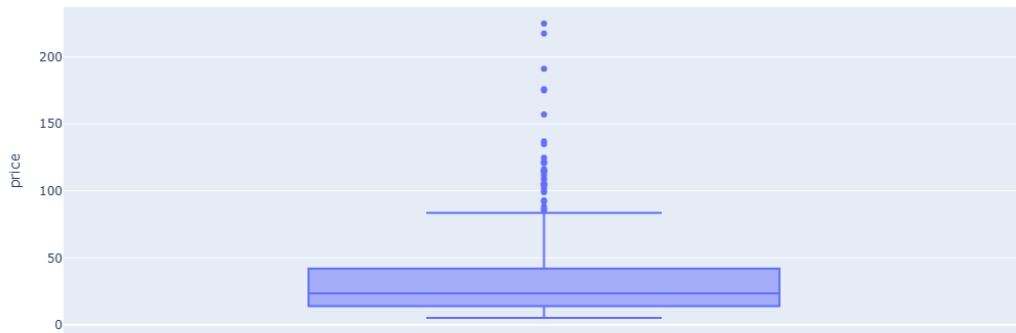
0.20 Utilisation de l'écart interquartile (IQR)

Cette fonction utilise la methode de l'écart intercatile

```
[374]: detection_des_outliers(erp_liaison_web, 'price')
```

```
dans la colonne price nous avons 31 outliers
```

Boxplot de price avec les outliers



```
[374]:
```

| | product_id | onsale_web | price | stock_quantity | stock_status | \ |
|-----|------------|------------|-------|----------------|--------------|---|
| 63 | 4115 | 1 | 100.0 | 12 | instock | |
| 65 | 4132 | 1 | 88.4 | 7 | instock | |
| 199 | 4352 | 1 | 225.0 | 0 | outofstock | |
| 205 | 4359 | 1 | 85.6 | 112 | instock | |
| 218 | 4402 | 1 | 176.0 | 11 | instock | |
| 219 | 4404 | 1 | 108.5 | 17 | instock | |
| 221 | 4406 | 1 | 157.0 | 12 | instock | |
| 222 | 4407 | 1 | 104.0 | 14 | instock | |
| 227 | 4582 | 1 | 109.6 | 18 | instock | |
| 380 | 4903 | 1 | 102.3 | 12 | instock | |
| 381 | 4904 | 1 | 137.0 | 9 | instock | |
| 426 | 5001 | 1 | 217.5 | 18 | instock | |
| 431 | 5007 | 1 | 105.0 | 15 | instock | |
| 432 | 5008 | 1 | 105.0 | 12 | instock | |
| 437 | 5025 | 1 | 112.0 | 136 | instock | |
| 438 | 5026 | 1 | 86.8 | 101 | instock | |
| 502 | 5565 | 1 | 92.0 | 0 | outofstock | |
| 511 | 5612 | 1 | 124.8 | 19 | instock | |
| 553 | 5767 | 1 | 175.0 | 12 | instock | |
| 587 | 5892 | 1 | 191.3 | 98 | instock | |
| 602 | 5916 | 1 | 93.0 | 1 | instock | |
| 603 | 5917 | 1 | 122.0 | 12 | instock | |
| 604 | 5918 | 1 | 114.0 | 12 | instock | |
| 642 | 6126 | 1 | 135.0 | 138 | instock | |
| 647 | 6201 | 1 | 105.6 | 16 | instock | |
| 648 | 6202 | 1 | 116.4 | 12 | instock | |
| 653 | 6212 | 1 | 115.0 | 16 | instock | |
| 654 | 6213 | 1 | 121.0 | 9 | instock | |
| 655 | 6214 | 1 | 99.0 | 9 | instock | |

| | | | | | |
|-----|------|---|-------|----|---------|
| 656 | 6215 | 1 | 115.0 | 14 | instock |
| 657 | 6216 | 1 | 121.0 | 14 | instock |

| | purchase_price | id_web | sku | virtual | downloadable | ... | \ |
|-----|----------------|--------|-------|---------|--------------|-----|---|
| 63 | 52.70 | 15382 | 15382 | 0 | 0 | ... | |
| 65 | 44.30 | 11668 | 11668 | 0 | 0 | ... | |
| 199 | 137.81 | 15940 | 15940 | 0 | 0 | ... | |
| 205 | 51.93 | 13853 | 13853 | 0 | 0 | ... | |
| 218 | 78.25 | 3510 | 3510 | 0 | 0 | ... | |
| 219 | 52.22 | 3507 | 3507 | 0 | 0 | ... | |
| 221 | 69.08 | 7819 | 7819 | 0 | 0 | ... | |
| 222 | 46.71 | 3509 | 3509 | 0 | 0 | ... | |
| 227 | 53.80 | 12857 | 12857 | 0 | 0 | ... | |
| 380 | 51.80 | 14805 | 14805 | 0 | 0 | ... | |
| 381 | 67.95 | 14220 | 14220 | 0 | 0 | ... | |
| 426 | 116.87 | 14581 | 14581 | 0 | 0 | ... | |
| 431 | 55.88 | 12791 | 12791 | 0 | 0 | ... | |
| 432 | 56.42 | 11602 | 11602 | 0 | 0 | ... | |
| 437 | 68.60 | 13914 | 13914 | 0 | 0 | ... | |
| 438 | 50.13 | 13913 | 13913 | 0 | 0 | ... | |
| 502 | 46.11 | 19822 | 19822 | 0 | 0 | ... | |
| 511 | 66.41 | 14915 | 14915 | 0 | 0 | ... | |
| 553 | 90.42 | 15185 | 15185 | 0 | 0 | ... | |
| 587 | 116.06 | 14983 | 14983 | 0 | 0 | ... | |
| 602 | 40.49 | 14774 | 14774 | 0 | 0 | ... | |
| 603 | 54.24 | 14775 | 14775 | 0 | 0 | ... | |
| 604 | 52.25 | 14773 | 14773 | 0 | 0 | ... | |
| 642 | 80.33 | 14923 | 14923 | 0 | 0 | ... | |
| 647 | 57.29 | 14596 | 14596 | 0 | 0 | ... | |
| 648 | 63.15 | 15126 | 15126 | 0 | 0 | ... | |
| 653 | 59.42 | 13996 | 13996 | 0 | 0 | ... | |
| 654 | 63.14 | 15072 | 15072 | 0 | 0 | ... | |
| 655 | 49.62 | 11601 | 11601 | 0 | 0 | ... | |
| 656 | 56.45 | 12790 | 12790 | 0 | 0 | ... | |
| 657 | 60.02 | 15070 | 15070 | 0 | 0 | ... | |

| | post_name | post_modified | \ |
|-----|---|---------------------|---|
| 63 | zind-humbrecht-riesling-gc-rangen-thann-clos-s... | 2020-02-08 11:45:02 | |
| 65 | zind-humbrecht-pinot-gris-grand-cru-rangen-de-... | 2020-02-20 09:55:02 | |
| 199 | champagne-egly-ouriet-grand-cru-millesime-2008 | 2020-03-07 11:18:45 | |
| 205 | champagne-larmandier-bernier-grand-cru-vieille... | 2019-12-23 09:30:11 | |
| 218 | cognac-frapin-vip-xo | 2020-08-22 11:35:03 | |
| 219 | cognac-frapin-fontpinot-xo | 2020-08-12 09:30:16 | |
| 221 | cognac-frapin-chateau-de-fontpinot-1989-20-ans | 2020-03-14 16:05:04 | |
| 222 | cognac-frapin-cigar-blend | 2020-07-04 09:45:03 | |
| 227 | chateau-de-puligny-montrachet-1cru-champ-canet... | 2020-02-06 16:35:02 | |
| 380 | domaine-des-croix-corton-grand-cru-les-greves-... | 2020-06-27 09:00:07 | |

| | | |
|-----|---|---------------------|
| 381 | domaine-des-croix-corton-charlemagne-grand-cru... | 2020-05-19 17:15:02 |
| 426 | david-duband-charmes-chambertin-grand-cru-2014 | 2020-05-16 09:00:05 |
| 431 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 2020-07-02 09:30:03 |
| 432 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 2020-06-23 15:35:02 |
| 437 | champagne-agrapart-fils-lavizoise-grand-cru-20... | 2020-07-09 17:05:02 |
| 438 | champagne-agrapart-fils-mineral-extra-brut-bla... | 2020-05-11 14:35:02 |
| 502 | tempier-bandol-cabassaou-2017 | 2020-01-04 13:57:04 |
| 511 | domaine-weinbach-gewurztraminer-gc-furstentum-... | 2019-01-23 09:33:57 |
| 553 | camille-giroud-clos-de-vougeot-2016 | 2020-06-11 15:25:04 |
| 587 | coteaux-champenois-egly-ouriet-ambonnay-rouge-... | 2020-04-01 09:30:09 |
| 602 | wemyss-malts-single-cask-chocolate-moka-cake | 2019-12-23 09:30:21 |
| 603 | wemyss-malts-single-cask-scotch-whisky-choc-n-... | 2020-03-11 09:30:09 |
| 604 | wemyss-malts-single-cask-scotch-whisky-chai-ca... | 2020-07-31 18:25:03 |
| 642 | champagne-gosset-celebris-vintage-2007 | 2020-08-27 11:45:02 |
| 647 | david-duband-chambolle-musigny-1er-cru-les-sen... | 2020-02-29 15:25:02 |
| 648 | domaine-clerget-echezeaux-en-orveaux-2015 | 2020-06-06 15:45:01 |
| 653 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 2020-06-16 09:30:16 |
| 654 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 2020-06-25 09:30:06 |
| 655 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 2020-07-04 11:35:02 |
| 656 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 2019-11-04 09:30:25 |
| 657 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 2020-07-30 09:30:08 |

| | post_modified_gmt | post_parent | \ |
|-----|---------------------|-------------|---|
| 63 | 2020-02-08 10:45:02 | 0.0 | |
| 65 | 2020-02-20 08:55:02 | 0.0 | |
| 199 | 2020-03-07 10:18:45 | 0.0 | |
| 205 | 2019-12-23 08:30:11 | 0.0 | |
| 218 | 2020-08-22 09:35:03 | 0.0 | |
| 219 | 2020-08-12 07:30:16 | 0.0 | |
| 221 | 2020-03-14 15:05:04 | 0.0 | |
| 222 | 2020-07-04 07:45:03 | 0.0 | |
| 227 | 2020-02-06 15:35:02 | 0.0 | |
| 380 | 2020-06-27 07:00:07 | 0.0 | |
| 381 | 2020-05-19 15:15:02 | 0.0 | |
| 426 | 2020-05-16 07:00:05 | 0.0 | |
| 431 | 2020-07-02 07:30:03 | 0.0 | |
| 432 | 2020-06-23 13:35:02 | 0.0 | |
| 437 | 2020-07-09 15:05:02 | 0.0 | |
| 438 | 2020-05-11 12:35:02 | 0.0 | |
| 502 | 2020-01-04 12:57:04 | 0.0 | |
| 511 | 2019-01-23 08:33:57 | 0.0 | |
| 553 | 2020-06-11 13:25:04 | 0.0 | |
| 587 | 2020-04-01 07:30:09 | 0.0 | |
| 602 | 2019-12-23 08:30:21 | 0.0 | |
| 603 | 2020-03-11 08:30:09 | 0.0 | |
| 604 | 2020-07-31 16:25:03 | 0.0 | |
| 642 | 2020-08-27 09:45:02 | 0.0 | |

| | | |
|-----|---------------------|-----|
| 647 | 2020-02-29 14:25:02 | 0.0 |
| 648 | 2020-06-06 13:45:01 | 0.0 |
| 653 | 2020-06-16 07:30:16 | 0.0 |
| 654 | 2020-06-25 07:30:06 | 0.0 |
| 655 | 2020-07-04 09:35:02 | 0.0 |
| 656 | 2019-11-04 08:30:25 | 0.0 |
| 657 | 2020-07-30 07:30:08 | 0.0 |

| | | guid | menu_order | post_type | \ |
|-----|---|------|------------|-----------|---|
| 63 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 65 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 199 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 205 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 218 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 219 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 221 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 222 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 227 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 380 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 381 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 426 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 431 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 432 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 437 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 438 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 502 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 511 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 553 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 587 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 602 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 603 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 604 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 642 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 647 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 648 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 653 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 654 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 655 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 656 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |
| 657 | https://www.bottle-neck.fr/?post_type=product&... | 0.0 | product | | |

| | comment_count | CA | z_score |
|-----|---------------|--------|----------|
| 63 | 0.0 | 100.0 | 2.453723 |
| 65 | 0.0 | 442.0 | 2.033082 |
| 199 | 0.0 | 2475.0 | 6.986486 |
| 205 | 0.0 | 599.2 | 1.931549 |
| 218 | 0.0 | 528.0 | 5.209643 |

| | | | |
|-----|-----|--------|----------|
| 219 | 0.0 | 434.0 | 2.761951 |
| 221 | 0.0 | 628.0 | 4.520663 |
| 222 | 0.0 | 520.0 | 2.598771 |
| 227 | 0.0 | 109.6 | 2.801839 |
| 380 | 0.0 | 204.6 | 2.537126 |
| 381 | 0.0 | 411.0 | 3.795421 |
| 426 | 0.0 | 435.0 | 6.714520 |
| 431 | 0.0 | 315.0 | 2.635033 |
| 432 | 0.0 | 735.0 | 2.635033 |
| 437 | 0.0 | 672.0 | 2.888868 |
| 438 | 0.0 | 781.2 | 1.975063 |
| 502 | 0.0 | 92.0 | 2.163626 |
| 511 | 0.0 | 124.8 | 3.353023 |
| 553 | 0.0 | 700.0 | 5.173380 |
| 587 | 0.0 | 1147.8 | 5.764453 |
| 602 | 0.0 | 93.0 | 2.199888 |
| 603 | 0.0 | 366.0 | 3.251489 |
| 604 | 0.0 | 342.0 | 2.961392 |
| 642 | 0.0 | 675.0 | 3.722896 |
| 647 | 0.0 | 633.6 | 2.656791 |
| 648 | 0.0 | 582.0 | 3.048421 |
| 653 | 0.0 | 805.0 | 2.997654 |
| 654 | 0.0 | 363.0 | 3.215227 |
| 655 | 0.0 | 594.0 | 2.417461 |
| 656 | 0.0 | 460.0 | 2.997654 |
| 657 | 0.0 | 242.0 | 3.215227 |

[31 rows x 33 columns]

0.20.1 Essayons maintenant avec la méthode des quartiles

```
[377]: Q1 = np.percentile(erp_liaison_web['price'],25)
print(f" Le premier quartile : {Q1:.1f} ")
```

Le premier quartile : 14.1

```
[379]: Q3 = np.percentile(erp_liaison_web['price'],75)
print(f" Le troisième quartile: {Q3:.1f} ")
```

Le troisième quartile: 42.1

```
[381]: IQR = Q3 - Q1

print(f" L'ecart inter quartile : {IQR:.1f} ")
```

L'ecart inter quartile : 28.0

```
[383]: la_borne_inf = Q1 - 1.5 * IQR
la_borne_sup = Q3 + 1.5 * IQR

print(f" la borne superieur : {la_borne_sup:.1f} , la borne inferieur : \u2192{la_borne_inf:.1f} ")
```

la borne superieur : 84.1 , la borne inferieur : -28.0

```
[385]: valeurs = erp_liaison_web[ ( erp_liaison_web['price'] < la_borne_inf ) | \u2192(erp_liaison_web['price'] > la_borne_sup) ]
print(f" Avec la m\u00e9thode de l'\u00e9cart interquartile on trouve {len(valeurs)}\u2192outliers ")
```

Avec la m\u00e9thode de l'\u00e9cart interquartile on trouve 31 outliers

```
[387]: OTVP = valeurs.sort_values(by='price', ascending = False)
OTVP[['product_id', 'post_name', 'price', 'purchase_price', 'z_score']]
```

```
[387]:      product_id      post_name  price \
199      4352  champagne-egly-ouriet-grand-cru-millesime-2008  225.0
426      5001  david-duband-charmes-chambertin-grand-cru-2014  217.5
587      5892  coteaux-champenois-egly-ouriet-ambonnay-rouge-...  191.3
218      4402                                cognac-frapin-vip-xo  176.0
553      5767                camille-giroud-clos-de-vougeot-2016  175.0
221      4406  cognac-frapin-chateau-de-fontpinot-1989-20-ans  157.0
381      4904  domaine-des-croix-corton-charlemagne-grand-cru-...  137.0
642      6126  champagne-gosset-celebris-vintage-2007  135.0
511      5612  domaine-weinbach-gewurztraminer-gc-furstentum-...  124.8
603      5917  wemyss-malts-single-cask-scotch-whisky-choc-n-...  122.0
654      6213  domaine-des-comtes-lafon-volnay-1er-cru-santen...  121.0
657      6216  domaine-des-comtes-lafon-volnay-1er-cru-champa...  121.0
648      6202  domaine-clerget-echezeaux-en-orveaux-2015  116.4
653      6212  domaine-des-comtes-lafon-volnay-1er-cru-santen...  115.0
656      6215  domaine-des-comtes-lafon-volnay-1er-cru-champa...  115.0
604      5918  wemyss-malts-single-cask-scotch-whisky-chai-ca...  114.0
437      5025  champagne-agrapart-fils-lavizoise-grand-cru-20...  112.0
227      4582  chateau-de-puligny-montrachet-1cru-champ-canet...  109.6
219      4404                                cognac-frapin-fontpinot-xo  108.5
647      6201  david-duband-chambolle-musigny-1er-cru-les-sen...  105.6
432      5008  domaine-des-comtes-lafon-volnay-1er-cru-santen...  105.0
431      5007  domaine-des-comtes-lafon-volnay-1er-cru-santen...  105.0
222      4407                                cognac-frapin-cigar-blend  104.0
380      4903  domaine-des-croix-corton-grand-cru-les-greves-...  102.3
63      4115  zind-humbrecht-riesling-gc-rangen-thann-clos-s...  100.0
655      6214  domaine-des-comtes-lafon-volnay-1er-cru-champa...  99.0
602      5916  wemyss-malts-single-cask-chocolate-moka-cake  93.0
502      5565                                tempier-bandol-cabassaou-2017  92.0
65      4132  zind-humbrecht-pinot-gris-grand-cru-rangen-de-...  88.4
```

| | | | |
|-----|------|---|------|
| 438 | 5026 | champagne-agrapart-fils-mineral-extra-brut-bla... | 86.8 |
| 205 | 4359 | champagne-larmandier-bernier-grand-cru-vieille... | 85.6 |

| | purchase_price | z_score |
|-----|----------------|----------|
| 199 | 137.81 | 6.986486 |
| 426 | 116.87 | 6.714520 |
| 587 | 116.06 | 5.764453 |
| 218 | 78.25 | 5.209643 |
| 553 | 90.42 | 5.173380 |
| 221 | 69.08 | 4.520663 |
| 381 | 67.95 | 3.795421 |
| 642 | 80.33 | 3.722896 |
| 511 | 66.41 | 3.353023 |
| 603 | 54.24 | 3.251489 |
| 654 | 63.14 | 3.215227 |
| 657 | 60.02 | 3.215227 |
| 648 | 63.15 | 3.048421 |
| 653 | 59.42 | 2.997654 |
| 656 | 56.45 | 2.997654 |
| 604 | 52.25 | 2.961392 |
| 437 | 68.60 | 2.888868 |
| 227 | 53.80 | 2.801839 |
| 219 | 52.22 | 2.761951 |
| 647 | 57.29 | 2.656791 |
| 432 | 56.42 | 2.635033 |
| 431 | 55.88 | 2.635033 |
| 222 | 46.71 | 2.598771 |
| 380 | 51.80 | 2.537126 |
| 63 | 52.70 | 2.453723 |
| 655 | 49.62 | 2.417461 |
| 602 | 40.49 | 2.199888 |
| 502 | 46.11 | 2.163626 |
| 65 | 44.30 | 2.033082 |
| 438 | 50.13 | 1.975063 |
| 205 | 51.93 | 1.931549 |

- **31 produits** ont été identifiés comme ayant un prix fortement au-dessus des bornes statistiques normales.
- Ces produits incluent principalement des **champagnes, cognacs, grands crus et spiritueux premium**, ce qui peut justifier leur écart de prix.
- Par exemple :
 - *Champagne Egly-Ouriet Grand Cru Millésime 2008* : 225 FCFA
 - *Cognac Frapin VIP XO* : 176 FCFA
 - *Domaine des Comtes Lafon Volnay 1er Cru* (plusieurs déclinaisons entre 105 FCFA et 121 FCFA)

Ces valeurs ne sont pas nécessairement des erreurs, mais elles doivent être **vérifiées manuellement** en lien avec la politique tarifaire. On peut aussi les **exclure temporairement**

lors d'analyses sensibles au prix (moyenne, cluster, corrélations).

ces prix atypiques peuvent refléter des produits rares ou une politique tarifaire non uniforme. On va vérifier si ces produits outre d'être cher sont vraiment rentables, et pour ça on va calculer la marge

[]:

0.20.2 Calcul du taux de marge, ce taux nous permet de voir si un produit est rentable

- si TVA = 20 %

```
[498]: erp_liaison_web['price_ht'] = erp_liaison_web['price'] / 1.2
erp_liaison_web['marge_unitaire'] = erp_liaison_web['price_ht'] -
↳ erp_liaison_web['purchase_price']
erp_liaison_web['taux_de_marge'] = erp_liaison_web['marge_unitaire'] /
↳ erp_liaison_web['price_ht']

[500]: erp_liaison_web['marge_totale'] = erp_liaison_web['marge_unitaire'] *
↳ erp_liaison_web['total_sales']
valeurs = erp_liaison_web[ ( erp_liaison_web['price'] < la_borne_inf ) |
↳ (erp_liaison_web['price'] > la_borne_sup) ]

marge_total = erp_liaison_web['marge_totale'].sum()

les_outliers_marge = valeurs['marge_unitaire'] * valeurs['total_sales']
marge_outliers = les_outliers_marge.sum()

la_part_marge_outliers = marge_outliers / marge_total

print(f"la marge totale : {marge_total:.2F} ")
print(f"la marge des outliers : {marge_outliers:.2F} ")
print(f"la part de marge des outliers : {la_part_marge_outliers:.2F} ")
```

```
la marge totale : 44660.65
la marge des outliers : 4873.46
la part de marge des outliers : 0.11
```

[513]:

```
[513]: Index(['product_id', 'onsale_web', 'price', 'stock_quantity', 'stock_status',
'purchase_price', 'id_web', 'sku', 'virtual', 'downloadable',
'rating_count', 'average_rating', 'total_sales', 'tax_status',
'post_author', 'post_date', 'post_date_gmt', 'product_type',
'post_title', 'post_excerpt', 'post_status', 'comment_status',
```

```
'ping_status', 'post_name', 'post_modified', 'post_modified_gmt',
'post_parent', 'guid', 'menu_order', 'post_type', 'comment_count', 'CA',
'z_score', 'price_ht', 'marge_unitaire', 'taux_de_marge',
'marge_totale', 'rotation_stock', 'duree_stock_mois',
'alerte_produit_anomalie_critique', 'alerte_stock_dormant'],
dtype='object')
```

0.20.3 Rotation des stocks

```
[400]: erp_liaison_web['rotation_stock'] = erp_liaison_web['total_sales'] /
↳(erp_liaison_web['stock_quantity'] + 1)
```

0.20.4 Durée moyenne des stocks en mois

```
[403]: erp_liaison_web['duree_stock_mois'] = 12 / (erp_liaison_web['rotation_stock'] +
↳1e-6)
```

```
[ ]:
```

0.20.5 Vérification : produits premium ou erreurs de tarification

Pour analyser les produits à prix extrêmes, nous avons croisé plusieurs critères métier :

- **Produit très cher + vendu souvent** → probablement un **produit premium** (positionnement volontaire)
- **Produit très cher + jamais vendu** → probablement une **erreur de prix** ou produit inadapté au marché
- **Produit très cher + marge faible ou négative** → tarification incohérente, à corriger
- **Produit très cher + bonne marge** → à conserver comme produit haut de gamme

```
[407]: valeurs = erp_liaison_web[ ( erp_liaison_web['price'] < la_borne_inf ) |
↳(erp_liaison_web['price'] > la_borne_sup) ]
valeurs[['product_id', 'post_name', 'price', 'purchase_price', 'marge_unitaire', 'total_sales', 'rot
↳sort_values(by='duree_stock_mois')]
```

```
[407]:
```

| | product_id | post_name | price \ |
|-----|------------|---|---------|
| 199 | 4352 | champagne-egly-ouriet-grand-cru-millesime-2008 | 225.0 |
| 502 | 5565 | tempier-bandol-cabassaou-2017 | 92.0 |
| 65 | 4132 | zind-humbrecht-pinot-gris-grand-cru-rangen-de-... | 88.4 |
| 655 | 6214 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 99.0 |
| 432 | 5008 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 105.0 |
| 602 | 5916 | wemyss-malts-single-cask-chocolate-moka-cake | 93.0 |
| 653 | 6212 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 115.0 |
| 648 | 6202 | domaine-clerget-echezeaux-en-orveaux-2015 | 116.4 |
| 647 | 6201 | david-duband-chambolle-musigny-1er-cru-les-sen... | 105.6 |
| 222 | 4407 | cognac-frapin-cigar-blend | 104.0 |
| 553 | 5767 | camille-giroud-clos-de-vougeot-2016 | 175.0 |
| 221 | 4406 | cognac-frapin-chateau-de-fontpinot-1989-20-ans | 157.0 |

| | | | |
|-----|------|---|-------|
| 381 | 4904 | domaine-des-croix-corton-charlemagne-grand-cru... | 137.0 |
| 654 | 6213 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 121.0 |
| 656 | 6215 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 115.0 |
| 218 | 4402 | cognac-frapin-vip-xo | 176.0 |
| 603 | 5917 | wemyss-malts-single-cask-scotch-whisky-choc-n-... | 122.0 |
| 604 | 5918 | wemyss-malts-single-cask-scotch-whisky-chai-ca... | 114.0 |
| 219 | 4404 | cognac-frapin-fontpinot-xo | 108.5 |
| 431 | 5007 | domaine-des-comtes-lafon-volnay-1er-cru-santen... | 105.0 |
| 380 | 4903 | domaine-des-croix-corton-grand-cru-les-greves-... | 102.3 |
| 657 | 6216 | domaine-des-comtes-lafon-volnay-1er-cru-champa... | 121.0 |
| 426 | 5001 | david-duband-charmes-chambertin-grand-cru-2014 | 217.5 |
| 438 | 5026 | champagne-agrapart-fils-mineral-extra-brut-bla... | 86.8 |
| 63 | 4115 | zind-humbrecht-riesling-gc-rangen-thann-clos-s... | 100.0 |
| 205 | 4359 | champagne-larmandier-bernier-grand-cru-vieille... | 85.6 |
| 587 | 5892 | coteaux-champenois-egly-ouriet-ambonnay-rouge-... | 191.3 |
| 227 | 4582 | chateau-de-puligny-montrachet-1cru-champ-canet... | 109.6 |
| 511 | 5612 | domaine-weinbach-gewurztraminer-gc-furstentum-... | 124.8 |
| 437 | 5025 | champagne-agrapart-fils-lavizoise-grand-cru-20... | 112.0 |
| 642 | 6126 | champagne-gosset-celebris-vintage-2007 | 135.0 |

| | purchase_price | marge_unitaire | total_sales | rotation_stock | \ |
|-----|----------------|----------------|-------------|----------------|---|
| 199 | 137.81 | 49.690000 | 11.0 | 11.000000 | |
| 502 | 46.11 | 30.556667 | 1.0 | 1.000000 | |
| 65 | 44.30 | 29.366667 | 5.0 | 0.625000 | |
| 655 | 49.62 | 32.880000 | 6.0 | 0.600000 | |
| 432 | 56.42 | 31.080000 | 7.0 | 0.538462 | |
| 602 | 40.49 | 37.010000 | 1.0 | 0.500000 | |
| 653 | 59.42 | 36.413333 | 7.0 | 0.411765 | |
| 648 | 63.15 | 33.850000 | 5.0 | 0.384615 | |
| 647 | 57.29 | 30.710000 | 6.0 | 0.352941 | |
| 222 | 46.71 | 39.956667 | 5.0 | 0.333333 | |
| 553 | 90.42 | 55.413333 | 4.0 | 0.307692 | |
| 221 | 69.08 | 61.753333 | 4.0 | 0.307692 | |
| 381 | 67.95 | 46.216667 | 3.0 | 0.300000 | |
| 654 | 63.14 | 37.693333 | 3.0 | 0.300000 | |
| 656 | 56.45 | 39.383333 | 4.0 | 0.266667 | |
| 218 | 78.25 | 68.416667 | 3.0 | 0.250000 | |
| 603 | 54.24 | 47.426667 | 3.0 | 0.230769 | |
| 604 | 52.25 | 42.750000 | 3.0 | 0.230769 | |
| 219 | 52.22 | 38.196667 | 4.0 | 0.222222 | |
| 431 | 55.88 | 31.620000 | 3.0 | 0.187500 | |
| 380 | 51.80 | 33.450000 | 2.0 | 0.153846 | |
| 657 | 60.02 | 40.813333 | 2.0 | 0.133333 | |
| 426 | 116.87 | 64.380000 | 2.0 | 0.105263 | |
| 438 | 50.13 | 22.203333 | 9.0 | 0.088235 | |
| 63 | 52.70 | 30.633333 | 1.0 | 0.076923 | |
| 205 | 51.93 | 19.403333 | 7.0 | 0.061947 | |

| | | | | |
|-----|--------|-----------|-----|----------|
| 587 | 116.06 | 43.356667 | 6.0 | 0.060606 |
| 227 | 53.80 | 37.533333 | 1.0 | 0.052632 |
| 511 | 66.41 | 37.590000 | 1.0 | 0.050000 |
| 437 | 68.60 | 24.733333 | 6.0 | 0.043796 |
| 642 | 80.33 | 32.170000 | 5.0 | 0.035971 |

| | duree_stock_mois | marge_totale | z_score |
|-----|------------------|--------------|----------|
| 199 | 1.090909 | 546.590000 | 6.986486 |
| 502 | 11.999988 | 30.556667 | 2.163626 |
| 65 | 19.199969 | 146.833333 | 2.033082 |
| 655 | 19.999967 | 197.280000 | 2.417461 |
| 432 | 22.285673 | 217.560000 | 2.635033 |
| 602 | 23.999952 | 37.010000 | 2.199888 |
| 653 | 29.142786 | 254.893333 | 2.997654 |
| 648 | 31.199919 | 169.250000 | 3.048421 |
| 647 | 33.999904 | 184.260000 | 2.656791 |
| 222 | 35.999892 | 199.783333 | 2.598771 |
| 553 | 38.999873 | 221.653333 | 5.173380 |
| 221 | 38.999873 | 247.013333 | 4.520663 |
| 381 | 39.999867 | 138.650000 | 3.795421 |
| 654 | 39.999867 | 113.080000 | 3.215227 |
| 656 | 44.999831 | 157.533333 | 2.997654 |
| 218 | 47.999808 | 205.250000 | 5.209643 |
| 603 | 51.999775 | 142.280000 | 3.251489 |
| 604 | 51.999775 | 128.250000 | 2.961392 |
| 219 | 53.999757 | 152.786667 | 2.761951 |
| 431 | 63.999659 | 94.860000 | 2.635033 |
| 380 | 77.999493 | 66.900000 | 2.537126 |
| 657 | 89.999325 | 81.626667 | 3.215227 |
| 426 | 113.998917 | 128.760000 | 6.714520 |
| 438 | 135.998459 | 199.830000 | 1.975063 |
| 63 | 155.997972 | 30.633333 | 2.453723 |
| 205 | 193.711159 | 135.823333 | 1.931549 |
| 587 | 197.996733 | 260.140000 | 5.764453 |
| 227 | 227.995668 | 37.533333 | 2.801839 |
| 511 | 239.995200 | 37.590000 | 3.353023 |
| 437 | 273.993744 | 148.400000 | 2.888868 |
| 642 | 333.590726 | 160.850000 | 3.722896 |

[]:

0.20.6 Analyse approfondie des produits à prix extrêmes

Nous avons extrait 31 produits dont le prix dépasse largement les bornes statistiques. Pour chacun, nous avons étudié : - La marge unitaire - Le volume de ventes - La rotation de stock - La durée moyenne de stockage (en mois) - La marge totale générée

Résultats :

- Certains produits très chers (ex : *champagne-egly-ouriet*) se vendent très bien avec un bon renouvellement de stock. Ce sont clairement des produits **premium justifiés**.
- D'autres produits, comme *tempier-bandol* ou *chateau-de-puligny*, ont été vendus **une seule fois** en plusieurs années, avec des durées de stockage dépassant **200 mois**. Ces produits posent un **risque de stock immobilisé**.
- Quelques produits très chers ont été vendus à **faible fréquence** mais génèrent une bonne marge : ils méritent une **surveillance** ou un **ajustement de prix léger**.

Recommandations :

- Vérifier si les prix des produits vendus **1 à 2 fois seulement** sont justifiés (positionnement premium vs erreur)
- Analyser les stocks avec une durée > 36 mois : il peut s'agir d'**anomalies de saisie** ou de **stocks oubliés**
- Mettre en place un **système d'alerte automatique** pour les produits :
 - à prix élevé
 - avec marge faible
 - et très faible rotation

[]:

```
[539]: # ===== Définition des seuils pour les problèmes de performance ---
SEUIL_ROTATION_FAIBLE = 0.2
SEUIL_DUREE_LONGUE = 36
SEUIL_TAUX_MARGE_CRITIQUE = 0.10
SEUIL_VENTES_CRITIQUES = 5

# ===== Condition d'Outlier de Prix ---
condition_prix_aberrant = (
    (erp_liaison_web['price'] < la_borne_inf) |
    (erp_liaison_web['price'] > la_borne_sup)
)

# ===== Condition de Problème de Performance ---
condition_probleme_performance = (
    (erp_liaison_web['rotation_stock'] < SEUIL_ROTATION_FAIBLE) |
    (erp_liaison_web['duree_stock_mois'] > SEUIL_DUREE_LONGUE) |
    (erp_liaison_web['taux_de_marge'] < SEUIL_TAUX_MARGE_CRITIQUE) |
    (erp_liaison_web['total_sales'] < SEUIL_VENTES_CRITIQUES)
)

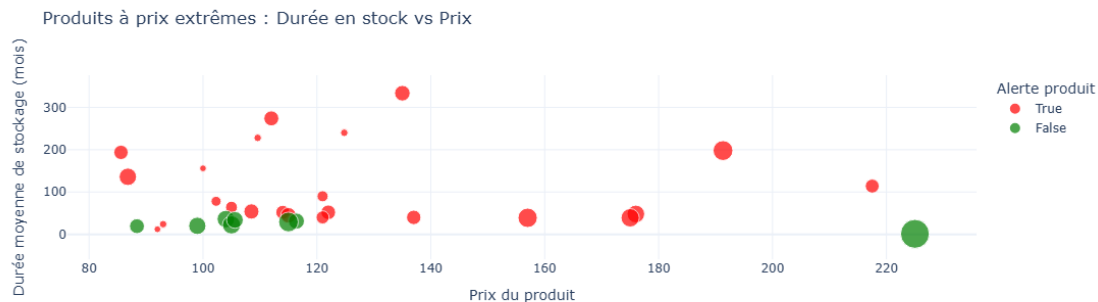
# ===== Combinaison des deux parties : Le produit est en alerte si
↳ c'est un outlier de prix ET qu'il a un problème de performance ---
erp_liaison_web['alerte_produit_anomalie_critique'] = (
    condition_prix_aberrant &
    condition_probleme_performance
)
```

```
[414]: valeurs = erp_liaison_web[ (erp_liaison_web['price'] < la_borne_inf) |
    ↪(erp_liaison_web['price'] > la_borne_sup) ]
```

```
[416]: fig = px.scatter(
    valeurs,
    x='price',
    y='duree_stock_mois',
    color='alerte_produit_anomalie_critique',
    size='marge_totale',
    hover_data=['post_name', 'total_sales', 'rotation_stock', 'marge_unitaire',
    ↪'marge_totale'],
    title='Produits à prix extrêmes : Durée en stock vs Prix',
    color_discrete_map={True: 'red', False: 'green'})

fig.update_layout(
    xaxis_title='Prix du produit ',
    yaxis_title='Durée moyenne de stockage (mois)',
    legend_title='Alerte produit',
    template='plotly_white'
)

fig.show()
```



```
[418]: fig = px.scatter(
    valeurs,
    x='price',
    y='total_sales',
    color='alerte_produit_anomalie_critique',
    size='marge_totale',
    hover_data=['post_name', 'total_sales', 'rotation_stock', 'marge_unitaire',
    ↪'marge_totale'],
    title='Produits à prix extrêmes : NBR Vente vs Prix',
```

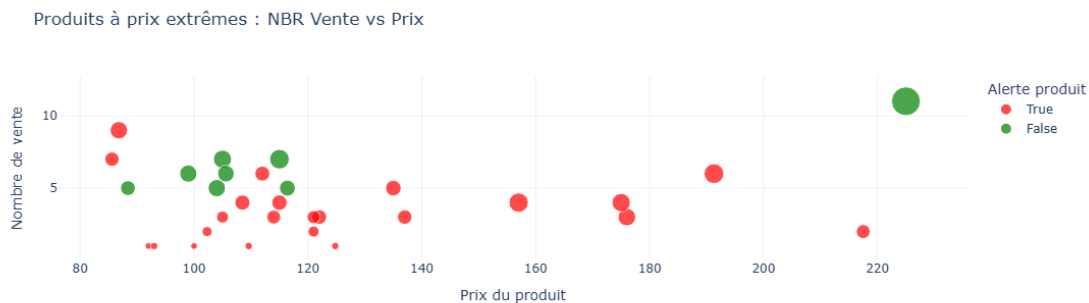
```

    color_discrete_map={True: 'red', False: 'green'}
)

fig.update_layout(
    xaxis_title='Prix du produit ',
    yaxis_title='Nombre de vente',
    legend_title='Alerte produit',
    template='plotly_white'
)

fig.show()

```



- Les produits avec anomalie

```

[421]: produits_anomalie_critique = □
        ↳ erp_liaison_web[erp_liaison_web['alerte_produit_anomalie_critique']]
        print(f"Nombre de produits identifiés comme 'Anomalie Critique' : □
        ↳ {len(produits_anomalie_critique)}")

```

Nombre de produits identifiés comme 'Anomalie Critique' : 23

```

[423]: produits_anomalie_critique[['post_title', 'price', 'total_sales', □
        ↳ 'stock_quantity', 'taux_de_marge', 'marge_totale', 'duree_stock_mois', □
        ↳ 'rotation_stock']]

```

```

[423]:

```

| | post_title | price | total_sales | \ |
|-----|---|-------|-------------|---|
| 63 | Zind-Humbrecht Riesling Grand Cru Rangen De Th... | 100.0 | 1.0 | |
| 205 | Champagne Larmandier-Bernier Grand Cru Les Che... | 85.6 | 7.0 | |
| 218 | Cognac Frapin VIP XO | 176.0 | 3.0 | |
| 219 | Cognac Frapin Château de Fontpinot XO | 108.5 | 4.0 | |
| 221 | Cognac Frapin Château de Fontpinot 1989 20 Ans... | 157.0 | 4.0 | |
| 227 | Château de Meursault Puligny-Montrachet 1er Cr... | 109.6 | 1.0 | |
| 380 | Domaine Des Croix Corton Grand Cru Les Grèves ... | 102.3 | 2.0 | |
| 381 | Domaine Des Croix Corton Charlemagne Grand Cru... | 137.0 | 3.0 | |
| 426 | David Duband Charmes-Chambertin Grand Cru 2014 | 217.5 | 2.0 | |

| | | | |
|-----|---|-------|-----|
| 431 | Domaine des Comtes Lafon Volnay 1er Cru Santen... | 105.0 | 3.0 |
| 437 | Champagne Agrapart & Fils L'Avizoise Extra... | 112.0 | 6.0 |
| 438 | Champagne Agrapart & Fils Minéral Extra Br... | 86.8 | 9.0 |
| 502 | Tempier Bandol Cabassaou 2017 | 92.0 | 1.0 |
| 511 | Domaine Weinbach Gewurztraminer Grand Cru Furs... | 124.8 | 1.0 |
| 553 | Camille Giroud Clos de Vougeot 2016 | 175.0 | 4.0 |
| 587 | Coteaux Champenois Egly-Ouriot Ambonnay Rouge ... | 191.3 | 6.0 |
| 602 | Wemyss Malts Single Cask Scotch Whisky Chocola... | 93.0 | 1.0 |
| 603 | Wemyss Malts Single Cask Scotch Whisky Choc 'n... | 122.0 | 3.0 |
| 604 | Wemyss Malts Single Cask Scotch Whisky Chai Ca... | 114.0 | 3.0 |
| 642 | Champagne Gosset Célébris Vintage 2007 | 135.0 | 5.0 |
| 654 | Domaine des Comtes Lafon Volnay 1er Cru Santen... | 121.0 | 3.0 |
| 656 | Domaine des Comtes Lafon Volnay 1er Cru Champa... | 115.0 | 4.0 |
| 657 | Domaine des Comtes Lafon Volnay 1er Cru Champa... | 121.0 | 2.0 |

| | stock_quantity | taux_de_marge | marge_totale | duree_stock_mois \ |
|-----|----------------|---------------|--------------|--------------------|
| 63 | 12 | 0.367600 | 30.633333 | 155.997972 |
| 205 | 112 | 0.272009 | 135.823333 | 193.711159 |
| 218 | 11 | 0.466477 | 205.250000 | 47.999808 |
| 219 | 17 | 0.422452 | 152.786667 | 53.999757 |
| 221 | 12 | 0.472000 | 247.013333 | 38.999873 |
| 227 | 18 | 0.410949 | 37.533333 | 227.995668 |
| 380 | 12 | 0.392375 | 66.900000 | 77.999493 |
| 381 | 9 | 0.404818 | 138.650000 | 39.999867 |
| 426 | 18 | 0.355200 | 128.760000 | 113.998917 |
| 431 | 15 | 0.361371 | 94.860000 | 63.999659 |
| 437 | 136 | 0.265000 | 148.400000 | 273.993744 |
| 438 | 101 | 0.306959 | 199.830000 | 135.998459 |
| 502 | 0 | 0.398565 | 30.556667 | 11.999988 |
| 511 | 19 | 0.361442 | 37.590000 | 239.995200 |
| 553 | 12 | 0.379977 | 221.653333 | 38.999873 |
| 587 | 98 | 0.271971 | 260.140000 | 197.996733 |
| 602 | 1 | 0.477548 | 37.010000 | 23.999952 |
| 603 | 12 | 0.466492 | 142.280000 | 51.999775 |
| 604 | 12 | 0.450000 | 128.250000 | 51.999775 |
| 642 | 138 | 0.285956 | 160.850000 | 333.590726 |
| 654 | 9 | 0.373818 | 113.080000 | 39.999867 |
| 656 | 14 | 0.410957 | 157.533333 | 44.999831 |
| 657 | 14 | 0.404760 | 81.626667 | 89.999325 |

| | rotation_stock |
|-----|----------------|
| 63 | 0.076923 |
| 205 | 0.061947 |
| 218 | 0.250000 |
| 219 | 0.222222 |
| 221 | 0.307692 |
| 227 | 0.052632 |

| | |
|-----|----------|
| 380 | 0.153846 |
| 381 | 0.300000 |
| 426 | 0.105263 |
| 431 | 0.187500 |
| 437 | 0.043796 |
| 438 | 0.088235 |
| 502 | 1.000000 |
| 511 | 0.050000 |
| 553 | 0.307692 |
| 587 | 0.060606 |
| 602 | 0.500000 |
| 603 | 0.230769 |
| 604 | 0.230769 |
| 642 | 0.035971 |
| 654 | 0.300000 |
| 656 | 0.266667 |
| 657 | 0.133333 |

- Ces produits sont identifiés comme présentant des anomalies critiques et sont classés comme ‘à risque’ pour plusieurs raisons. Bien que certains d’entre eux puissent être des produits haut de gamme dont le prix est naturellement élevé, leur classification en ‘risque’ découle d’une combinaison de facteurs :

Leur prix de vente se situe en dehors des bornes statistiques (définies par l’IQR), ce qui peut indiquer une erreur de saisie, un positionnement tarifaire incohérent, ou une valeur atypique nécessitant vérification. **Un prix très élevé, même justifié, peut par exemple limiter les ventes.**

Problème de Performance sous-jacent : En plus de cette anomalie de prix, ces produits présentent au moins un des indicateurs de performance préoccupants suivants :

Rotation de stock très faible : Le produit se vend très peu par rapport au stock disponible, ce qui ralentit la circulation du capital.

Durée de stock excessive : Le produit est immobilisé dans l’entrepôt depuis une période trop longue (plus de trois ans dans notre cas), représentant un coût de stockage et un risque d’obsolescence.

Marge critique ou trop faible : Le produit génère une marge insuffisante (inférieure à 10%) ou potentiellement nulle/négative, le rendant peu rentable, voire coûteux à vendre.

Ventes très faibles : Un volume de ventes minimal (moins de 5 unités sur la période) indique un manque d’attractivité ou de visibilité du produit.

[]:

0.21 Marge totale

Nous identifions ici les 10 produits les plus rentables selon la marge totale. Cela permet de repérer ceux qui génèrent le plus de profit unitaire multiplié par les ventes.

1 Top 10 des produits les plus rentables

```
[430]: top_marge = erp_liaison_web.sort_values(by='marge_totale', ascending=False).
        ↪head(10)
top_marge[['post_title', 'price', 'purchase_price', 'marge_unitaire',
        ↪'taux_de_marge', 'marge_totale', 'total_sales', 'rotation_stock', 'duree_stock_mois']]
```

```
[430]:
```

| | post_title | price | purchase_price | \ |
|-----|---|-------|----------------|---|
| 199 | Champagne Egly-Ouriet Grand Cru Millésimé 2008 | 225.0 | 137.81 | |
| 200 | Champagne Egly-Ouriet Grand Cru Brut Rosé | 79.5 | 45.91 | |
| 587 | Coteaux Champenois Egly-Ouriet Ambonnay Rouge ... | 191.3 | 116.06 | |
| 653 | Domaine des Comtes Lafon Volnay 1er Cru Santen... | 115.0 | 59.42 | |
| 582 | Agnès Levet Côte Rôtie Améthyste 2017 | 41.2 | 21.71 | |
| 221 | Cognac Frapin Château de Fontpinot 1989 20 Ans... | 157.0 | 69.08 | |
| 553 | Camille Giroud Clos de Vougeot 2016 | 175.0 | 90.42 | |
| 432 | Domaine des Comtes Lafon Volnay 1er Cru Santen... | 105.0 | 56.42 | |
| 265 | Bernard Baudry Chinon Rouge La Croix Boissée 2017 | 28.5 | 14.14 | |
| 218 | Cognac Frapin VIP XO | 176.0 | 78.25 | |

| | marge_unitaire | taux_de_marge | marge_totale | total_sales | rotation_stock | \ |
|-----|----------------|---------------|--------------|-------------|----------------|---|
| 199 | 49.690000 | 0.265013 | 546.590000 | 11.0 | 11.000000 | |
| 200 | 20.340000 | 0.307019 | 284.760000 | 14.0 | 0.109375 | |
| 587 | 43.356667 | 0.271971 | 260.140000 | 6.0 | 0.060606 | |
| 653 | 36.413333 | 0.379965 | 254.893333 | 7.0 | 0.411765 | |
| 582 | 12.623333 | 0.367670 | 252.466667 | 20.0 | 0.571429 | |
| 221 | 61.753333 | 0.472000 | 247.013333 | 4.0 | 0.307692 | |
| 553 | 55.413333 | 0.379977 | 221.653333 | 4.0 | 0.307692 | |
| 432 | 31.080000 | 0.355200 | 217.560000 | 7.0 | 0.538462 | |
| 265 | 9.610000 | 0.404632 | 211.420000 | 22.0 | 0.478261 | |
| 218 | 68.416667 | 0.466477 | 205.250000 | 3.0 | 0.250000 | |

| | duree_stock_mois |
|-----|------------------|
| 199 | 1.090909 |
| 200 | 109.713283 |
| 587 | 197.996733 |
| 653 | 29.142786 |
| 582 | 20.999963 |
| 221 | 38.999873 |
| 553 | 38.999873 |
| 432 | 22.285673 |
| 265 | 25.090857 |
| 218 | 47.999808 |

1.0.1 Produits les plus rentables en valeur (Top marge totale)

Nous avons extrait les 10 produits qui génèrent le plus de **marge totale**, c'est-à-dire ceux qui contribuent le plus directement à la rentabilité globale de l'entreprise.

- Tous ces produits ont une marge unitaire élevée (de 9 à 68 FCFA).
- Ils affichent aussi un taux de marge supérieur à 25 %, ce qui en fait des produits solides.
- Certains, comme le *Champagne Egly-Ouriel Millésimé 2008*, combinent à la fois **bon volume de vente** et **très bon positionnement tarifaire**.
- D'autres, comme les Cognacs Frapin, ont une rotation lente, mais génèrent une forte marge par unité → ce sont des produits **premium bien positionnés**.

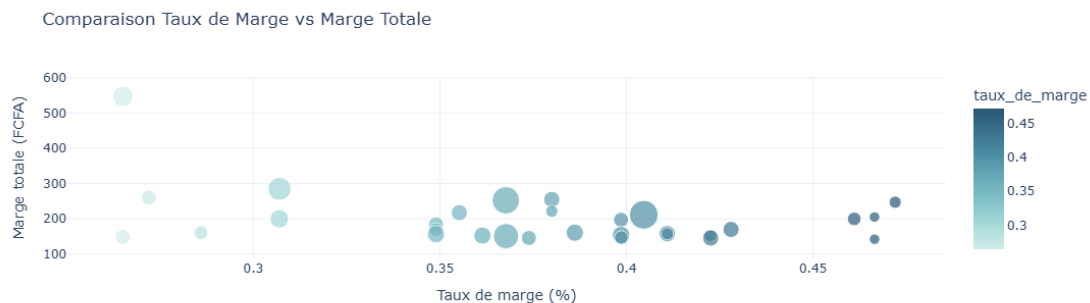
Ces produits doivent être considérés comme des **leviers prioritaires** pour la stratégie commerciale.

```
[433]: top_marge_plot = erp_liaison_web.sort_values(by='marge_totale',
↪ascending=False).head(30)

fig = px.scatter(
    top_marge_plot,
    x='taux_de_marge',
    y='marge_totale',
    size='total_sales',
    color='taux_de_marge',
    hover_data=['post_title', 'price', 'marge_unitaire', 'total_sales'],
    title='Comparaison Taux de Marge vs Marge Totale',
    color_continuous_scale='Teal'
)

fig.update_layout(
    xaxis_title='Taux de marge (%)',
    yaxis_title='Marge totale (FCFA)',
    template='plotly_white'
)

fig.show()
```



1.0.2 Visualisation croisée : Taux de Marge vs Marge Totale

Le graphique ci-dessous permet d'identifier visuellement les produits les plus performants en croisant : - leur **taux de marge (%)** → rentabilité unitaire - leur **marge totale (FCFA)** → contribution

réelle au profit

Les produits en haut à droite cumulent un bon taux de marge et un bon volume de ventes : ce sont les produits **les plus stratégiques** à conserver et à pousser commercialement.

[]:

[]:

2 Top 10 des produits les moins rentables

```
[439]: bottom_marge = erp_liaison_web.sort_values(by='marge_totale', ascending=True).
        ↪head(10)
        bottom_marge[['post_title', 'price', 'purchase_price', 'marge_unitaire',
        ↪'taux_de_marge', 'marge_totale', 'total_sales', 'rotation_stock', 'duree_stock_mois']]
```

```
[439]:
```

| | | post_title | price | purchase_price | \ |
|-----|--|---|-------|----------------|---|
| 116 | | Château de La Liquière Faugères Tucade 2015 | 27.2 | 14.05 | |
| 454 | | Liban Vallée de la Békaa Château Marsyas 2012 | 28.8 | 15.18 | |
| 248 | | Alphonse Mellot Sancerre Rouge La Moussière 2014 | 30.6 | 16.44 | |
| 41 | | Le Vieux Donjon Châteauneuf-du-Pape 2013 | 37.0 | 19.50 | |
| 212 | | Wemyss Malts Blended Malt Scotch Whisky The Hi... | 62.0 | 28.42 | |
| 367 | | Tempier Bandol Rosé 2019 | 22.8 | 11.90 | |
| 622 | | I Fabbri Chianti Classico Gran Selezione 2015 | 48.5 | 25.31 | |
| 54 | | Emile Boeckel Gewurztraminer Grand Cru Zotzenb... | 15.8 | 8.57 | |
| 16 | | Clos du Mont-Olivet Châteauneuf-du-Pape Blanc ... | 33.7 | 18.11 | |
| 15 | | Jeanne Gaillard IGP Collines Rhodaniennes Syra... | 7.7 | 4.14 | |

| | marge_unitaire | taux_de_marge | marge_totale | total_sales | rotation_stock | \ |
|-----|----------------|---------------|--------------|-------------|----------------|---|
| 116 | 8.616667 | 0.380147 | 0.0 | 0.0 | 0.0 | |
| 454 | 8.820000 | 0.367500 | 0.0 | 0.0 | 0.0 | |
| 248 | 9.060000 | 0.355294 | 0.0 | 0.0 | 0.0 | |
| 41 | 11.333333 | 0.367568 | 0.0 | 0.0 | 0.0 | |
| 212 | 23.246667 | 0.449935 | 0.0 | 0.0 | 0.0 | |
| 367 | 7.100000 | 0.373684 | 0.0 | 0.0 | 0.0 | |
| 622 | 15.106667 | 0.373773 | 0.0 | 0.0 | 0.0 | |
| 54 | 4.596667 | 0.349114 | 0.0 | 0.0 | 0.0 | |
| 16 | 9.973333 | 0.355134 | 0.0 | 0.0 | 0.0 | |
| 15 | 2.276667 | 0.354805 | 0.0 | 0.0 | 0.0 | |

| | duree_stock_mois |
|-----|------------------|
| 116 | 12000000.0 |
| 454 | 12000000.0 |
| 248 | 12000000.0 |
| 41 | 12000000.0 |
| 212 | 12000000.0 |
| 367 | 12000000.0 |

| | |
|-----|------------|
| 622 | 12000000.0 |
| 54 | 12000000.0 |
| 16 | 12000000.0 |
| 15 | 12000000.0 |

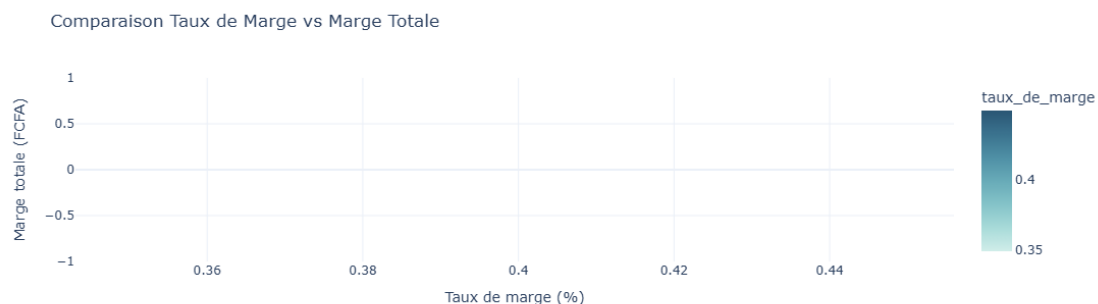
2.1 Comparaison taux de marge vs marge totale

```
[442]: bottom_marge = erp_liaison_web.sort_values(by='marge_totale', ascending=True).
        ↪head(10)

fig = px.scatter(
    bottom_marge,
    x='taux_de_marge',
    y='marge_totale',
    size='total_sales',
    color='taux_de_marge',
    hover_data=['post_title', 'price', 'marge_unitaire', 'total_sales'],
    title='Comparaison Taux de Marge vs Marge Totale',
    color_continuous_scale='Teal'
)

fig.update_layout(
    xaxis_title='Taux de marge (%)',
    yaxis_title='Marge totale (FCFA)',
    template='plotly_white'
)

fig.show()
```



Ces produits n'ont généré **aucune marge totale** car ils **n'ont jamais été vendus**. Pourtant, leur taux de marge est correct, ce qui indique un problème potentiel ailleurs :

- Produit bloqué ou indisponible en stock
- Erreur de publication sur le site

- Produit peu visible ou non attractif pour les clients

Ces cas doivent faire l'objet d'une **vérification commerciale ou marketing**. Un produit non vendu est un **manque à gagner potentiel**, surtout quand la marge est bonne.

[]:

[]:

[448]: `erp_liaison_web[erp_liaison_web['stock_quantity'] < 0]`

```
[448]:      product_id  onsale_web  price  stock_quantity  stock_status  \
528           5700           1   44.5             -1    outofstock

      purchase_price  id_web    sku  virtual  downloadable  ...  comment_count  \
528           22.3  14736  14736         0             0  ...             0.0

      CA    z_score  price_ht  marge_unitaire  taux_de_marge  marge_totale  \
528  356.0  0.441176  37.083333         14.783333         0.398652    118.266667

      rotation_stock  duree_stock_mois  alerte_produit_anomalie_critique
528              inf              0.0                          False

[1 rows x 40 columns]
```

Anomalie détectée

Un contrôle qualité des stocks a révélé une anomalie sur certains produits, dont le champ `stock_quantity` est négatif. Ce cas particulier (`stock = -1`) génère une durée de stockage infinie (`inf`) lors du calcul automatique. Cette valeur a été traitée comme une erreur de saisie dans le visuel précédent et exclue des visualisations et calculs agrégés, afin de ne pas biaiser l'analyse.

[]:

2.2 Rotation de stock

3 Top 10 des produits qui tournent le plus vite

```
[517]: erp_rotation_valide = erp_liaison_web[erp_liaison_web['stock_quantity'] >= 0]
top_rotation = erp_rotation_valide.sort_values(by='rotation_stock',
↪ascending=False).head(10)
top_rotation[['post_title', 'price', 'total_sales', 'stock_quantity',
↪'rotation_stock', 'duree_stock_mois']]
```

```
[517]:      post_title  price  total_sales  \
321  François Baur Pinot Noir Schlittweg 2017    12.7         22.0
366      Triennes IGP Méditerranée Rosé 2019     9.3         16.0
149  Parcé Frères IGP Côtes Catalanes Hommage à Fer...    8.9         15.0
607  Borie La Vitarèle Languedoc Blanc Le Grand May...   14.1         14.0
```

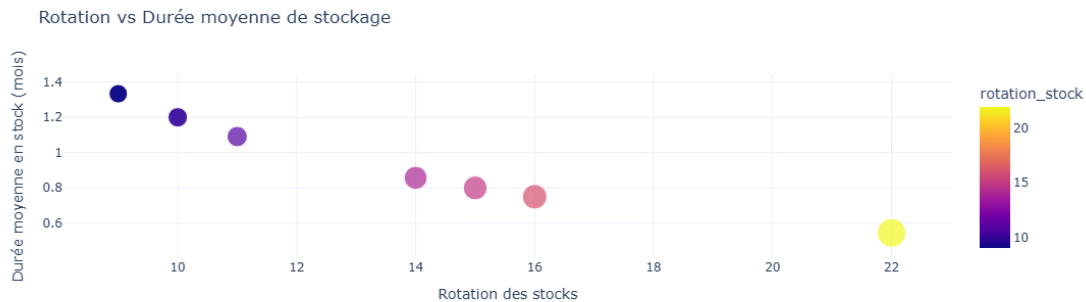
| | | | |
|-----|---|-------|------|
| 199 | Champagne Egly-Ouriét Grand Cru Millésimé 2008 | 225.0 | 11.0 |
| 368 | Saumaize-Michelin Mâcon Vergisson Sur La Roche... | 18.7 | 10.0 |
| 304 | Albert Mann Muscat 2018 | 16.8 | 10.0 |
| 516 | Albert Boxler Pinot Gris Réserve 2016 | 27.8 | 9.0 |
| 159 | Domaine Huet Vouvray Haut-Lieu Sec 2017 | 24.4 | 9.0 |
| 479 | Decelle-Villa Nuits-Saint-Georges Rouge 2016 | 43.5 | 9.0 |

| | stock_quantity | rotation_stock | duree_stock_mois |
|-----|----------------|----------------|------------------|
| 321 | 0 | 22.0 | 0.545455 |
| 366 | 0 | 16.0 | 0.750000 |
| 149 | 0 | 15.0 | 0.800000 |
| 607 | 0 | 14.0 | 0.857143 |
| 199 | 0 | 11.0 | 1.090909 |
| 368 | 0 | 10.0 | 1.200000 |
| 304 | 0 | 10.0 | 1.200000 |
| 516 | 0 | 9.0 | 1.333333 |
| 159 | 0 | 9.0 | 1.333333 |
| 479 | 0 | 9.0 | 1.333333 |

```
[519]: fig = px.scatter(
    top_rotation,
    x='rotation_stock',
    y='duree_stock_mois',
    size='total_sales',
    color='rotation_stock',
    hover_data=['post_title', 'total_sales', 'stock_quantity'],
    title="Rotation vs Durée moyenne de stockage",
    color_continuous_scale='Plasma'
)

fig.update_layout(
    xaxis_title='Rotation des stocks',
    yaxis_title='Durée moyenne en stock (mois)',
    template='plotly_white'
)

fig.show()
```



Ces produits affichent une rotation très élevée, signe d'une **forte demande client** et d'un **bon positionnement commercial** :

- Tous ont été **totalelement vendus** (stock = 0)
- Leur durée moyenne de stockage est **inférieure à 1,5 mois**

Ce sont des produits à fort potentiel qu'il convient de : - Recommander plus largement - **Sécuriser en stock** pour éviter les ruptures - Intégrer dans une **stratégie commerciale prioritaire**

[]:

[]:

```
[534]: top_10_fast_selling = erp_liaison_web.sort_values(by='rotation_stock',
    ↪ascending=False).head(10)

ca_top_10 = top_10_fast_selling['CA'].sum()

ca_total_global = erp_liaison_web['CA'].sum()

if ca_total_global > 0:
    part_ca_top_10 = (ca_top_10 / ca_total_global) * 100
else:
    part_ca_top_10 = 0

print(f"Chiffre d'Affaires total des 10 produits avec un bon taux de rotation :
    ↪{ca_top_10:,.2f} FCFA".format().replace(',', ' '))
print(f"Chiffre d'Affaires total de tous les produits : {ca_total_global:,.2f}
    ↪FCFA".format().replace(',', ' '))
print(f"Les 10 produits représentent {part_ca_top_10:,.2f} % du Chiffre
    ↪d'Affaires total.".format().replace(',', ' '))
```

Chiffre d'Affaires total des 10 produits avec un bon taux de rotation : 4 586.80
FCFA

Chiffre d'Affaires total de tous les produits : 143 680.10 FCFA
 Les 10 produits représentent 3.19 % du Chiffre d'Affaires total.

Ces produits se vendent à une très bonne vitesse mais ne représentent que 3% du CA total parce que la majorité des produits vendus à de très grande vitesse sont à bas prix

[]:

[]:

4 Top 10 des produits qui stagnent (rotation la plus lente)

```
[466]: bottom_rotation = erp_rotation_valide[erp_rotation_valide['rotation_stock'] < 100].sort_values(by='rotation_stock', ascending=True).head(10)
bottom_rotation[['post_title', 'total_sales', 'stock_quantity', 'rotation_stock', 'duree_stock_mois']]
```

```
[466]:
```

| | post_title | total_sales | \ |
|-----|---|-------------|---|
| 668 | Borie La Vitarèle Saint-Chinian Midi Rouge 2015 | 0.0 | |
| 67 | Zind-Humbrecht Zind 2017 | 0.0 | |
| 608 | Cognac Normandin Mercier VFC | 0.0 | |
| 201 | Champagne Egly-Ouriot Grand Cru Blanc de Noirs | 0.0 | |
| 622 | I Fabbri Chianti Classico Gran Selezione 2015 | 0.0 | |
| 212 | Wemyss Malts Blended Malt Scotch Whisky The Hi... | 0.0 | |
| 248 | Alphonse Mellot Sancerre Rouge La Moussière 2014 | 0.0 | |
| 41 | Le Vieux Donjon Châteauneuf-du-Pape 2013 | 0.0 | |
| 54 | Emile Boeckel Gewurztraminer Grand Cru Zotzenb... | 0.0 | |
| 16 | Clos du Mont-Olivet Châteauneuf-du-Pape Blanc ... | 0.0 | |

| | stock_quantity | rotation_stock | duree_stock_mois |
|-----|----------------|----------------|------------------|
| 668 | 0 | 0.0 | 12000000.0 |
| 67 | 0 | 0.0 | 12000000.0 |
| 608 | 13 | 0.0 | 12000000.0 |
| 201 | 97 | 0.0 | 12000000.0 |
| 622 | 0 | 0.0 | 12000000.0 |
| 212 | 0 | 0.0 | 12000000.0 |
| 248 | 0 | 0.0 | 12000000.0 |
| 41 | 0 | 0.0 | 12000000.0 |
| 54 | 0 | 0.0 | 12000000.0 |
| 16 | 0 | 0.0 | 12000000.0 |

Ces produits n'ont enregistré **aucune vente** depuis leur mise en base. Ils sont identifiés comme **produits dormants** avec une rotation nulle :

Caractéristiques : - **total_sales** = 0 → aucun client n'a acheté ces produits - **stock_quantity** = 0 ou → certains sont en stock, d'autres non - **duree_stock_mois** = 1 200 000 → durée fictive liée à l'absence de ventes

4.0.1 Hypothèses à vérifier :

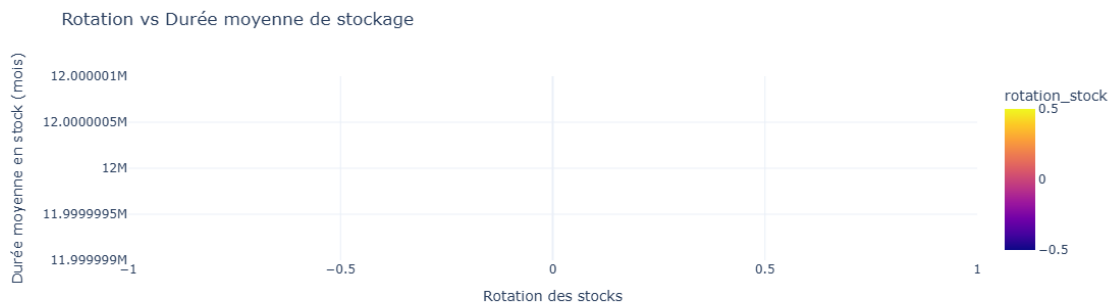
- Erreur de saisie (produit non activé sur le site)
- Produit mal référencé (erreur sur le `sku`, `id_web`)
- Produit non attractif (prix, photo, description)
- Produit jamais approvisionné ou épuisé depuis longtemps

4.0.2 Recommandations :

- Faire un audit de ces produits avec l'équipe métier
- Retirer ou corriger ceux qui ne seront jamais vendus
- Mettre en avant ceux qui pourraient redevenir actifs

Ces produits immobilisent inutilement des lignes en base ou du stock → **à nettoyer ou réactiver**.

```
[469]: fig = px.scatter(  
    bottom_rotation,  
    x='rotation_stock',  
    y='duree_stock_mois',  
    size='total_sales',  
    color='rotation_stock',  
    hover_data=['post_title', 'total_sales', 'stock_quantity'],  
    title="Rotation vs Durée moyenne de stockage",  
    color_continuous_scale='Plasma'  
)  
  
fig.update_layout(  
    xaxis_title='Rotation des stocks',  
    yaxis_title='Durée moyenne en stock (mois)',  
    template='plotly_white'  
)  
  
fig.show()
```



```
[ ]:
```

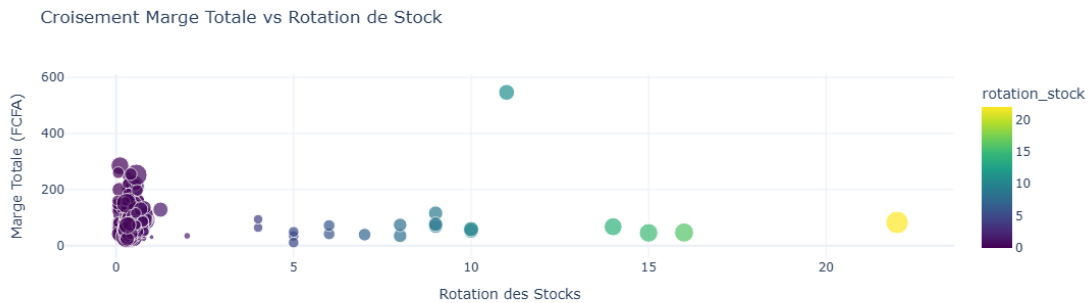


```
[537]: produits_valides = erp_liaison_web[erp_liaison_web['stock_quantity'] >= 0]

fig = px.scatter(
    produits_valides,
    x='rotation_stock',
    y='marge_totale',
    size='total_sales',
    color='rotation_stock',
    hover_data=['post_title', 'price', 'stock_quantity', 'marge_totale'],
    title='Croisement Marge Totale vs Rotation de Stock',
    color_continuous_scale='Viridis'
)

fig.update_layout(
    xaxis_title='Rotation des Stocks',
    yaxis_title='Marge Totale (FCFA)',
    template='plotly_white'
)

fig.show()
```



[]:

[]:

4.1 Produits dormants ou bloqués

Les produits suivants cumulent une très longue durée de stockage et peu de ventes, ce qui indique un risque d'immobilisation de capital ou un désintérêt client.

[474]:

```

produits_dormants = erp_liaison_web[(erp_liaison_web['duree_stock_mois'] > 100) &
    (erp_liaison_web['total_sales'] < 5)].sort_values(by='duree_stock_mois',
    ascending=False)
produits_dormants[['post_title', 'price', 'CA', 'purchase_price',
    'marge_unitaire',
    'taux_de_marge', 'rotation_stock', 'duree_stock_mois', 'stock_quantity', 'stock_status']]

```

[474]:

| | post_title | price | CA \ |
|-----|---|--------|-------|
| 2 | Pierre Jean Villa Crozes-Hermitage Accroche Co... | 20.80 | 0.0 |
| 196 | Champagne Mailly Grand Cru Les Echansons 2007 | 83.00 | 0.0 |
| 622 | I Fabbri Chianti Classico Gran Selezione 2015 | 48.50 | 0.0 |
| 608 | Cognac Normandin Mercier VFC | 59.90 | 0.0 |
| 536 | Albert Mann Gewurztraminer Vendanges Tardives ... | 57.60 | 0.0 |
| 533 | Domaine des Terres d'Ocre Saint-Pourçain Rouge... | 10.80 | 0.0 |
| 490 | Marc Colin Et Fils Chassagne-Montrachet Blanc ... | 61.60 | 0.0 |
| 454 | Liban Vallée de la Békaa Château Marsyas 2012 | 28.80 | 0.0 |
| 393 | Cave de Castelmaure Corbières Rouge Vigneron 2018 | 7.00 | 0.0 |
| 367 | Tempier Bandol Rosé 2019 | 22.80 | 0.0 |
| 248 | Alphonse Mellot Sancerre Rouge La Moussière 2014 | 30.60 | 0.0 |
| 237 | Catherine et Claude Maréchal Volnay 2017 | 50.10 | 0.0 |
| 212 | Wemyss Malts Blended Malt Scotch Whisky The Hi... | 62.00 | 0.0 |
| 8 | Pierre Gaillard Côte Rôtie Esprit de Blonde 2017 | 60.00 | 0.0 |
| 201 | Champagne Egly-Ouriet Grand Cru Blanc de Noirs | 12.65 | 0.0 |
| 118 | Moulin de Gassac IGP Pays d'Hérault Guilhem Ro... | 5.80 | 0.0 |
| 116 | Château de La Liquière Faugères Tucade 2015 | 27.20 | 0.0 |
| 11 | Pierre Gaillard Côtes-du-Rhône Blanc Les Gendr... | 18.30 | 0.0 |
| 15 | Jeanne Gaillard IGP Collines Rhodaniennes Syra... | 7.70 | 0.0 |
| 16 | Clos du Mont-Olivet Châteauneuf-du-Pape Blanc ... | 33.70 | 0.0 |
| 27 | Oratoire Saint Martin Cairanne Rouge Les Douye... | 19.50 | 0.0 |
| 41 | Le Vieux Donjon Châteauneuf-du-Pape 2013 | 37.00 | 0.0 |
| 54 | Emile Boeckel Gewurztraminer Grand Cru Zotzenb... | 15.80 | 0.0 |
| 668 | Borie La Vitarèle Saint-Chinian Midi Rouge 2015 | 40.50 | 0.0 |
| 67 | Zind-Humbrecht Zind 2017 | 25.70 | 0.0 |
| 70 | Champagne Gosset Grand Millésime 2006 | 53.00 | 212.0 |
| 202 | Champagne Egly-Ouriet Premier Cru Les Vignes d... | 51.60 | 154.8 |
| 74 | Champagne Mailly Grand Cru Brut Rosé | 37.50 | 112.5 |
| 71 | Champagne Gosset Grand Rosé | 49.00 | 196.0 |
| 511 | Domaine Weinbach Gewurztraminer Grand Cru Furs... | 124.80 | 124.8 |
| 227 | Château de Meursault Puligny-Montrachet 1er Cr... | 109.60 | 109.6 |
| 63 | Zind-Humbrecht Riesling Grand Cru Rangen De Th... | 100.00 | 100.0 |
| 426 | David Duband Charmes-Chambertin Grand Cru 2014 | 217.50 | 435.0 |

| | purchase_price | marge_unitaire | taux_de_marge | rotation_stock \ |
|-----|----------------|----------------|---------------|------------------|
| 2 | 10.64 | 6.693333 | 0.386154 | 0.000000 |
| 196 | 48.90 | 20.266667 | 0.293012 | 0.000000 |
| 622 | 25.31 | 15.106667 | 0.373773 | 0.000000 |
| 608 | 27.18 | 22.736667 | 0.455492 | 0.000000 |

| | | | | |
|-----|--------|------------|-----------|----------|
| 536 | 30.36 | 17.640000 | 0.367500 | 0.000000 |
| 533 | 5.69 | 3.310000 | 0.367778 | 0.000000 |
| 490 | 31.51 | 19.823333 | 0.386169 | 0.000000 |
| 454 | 15.18 | 8.820000 | 0.367500 | 0.000000 |
| 393 | 3.65 | 2.183333 | 0.374286 | 0.000000 |
| 367 | 11.90 | 7.100000 | 0.373684 | 0.000000 |
| 248 | 16.44 | 9.060000 | 0.355294 | 0.000000 |
| 237 | 24.59 | 17.160000 | 0.411018 | 0.000000 |
| 212 | 28.42 | 23.246667 | 0.449935 | 0.000000 |
| 8 | 29.45 | 20.550000 | 0.411000 | 0.000000 |
| 201 | 77.48 | -66.938333 | -6.349881 | 0.000000 |
| 118 | 2.97 | 1.863333 | 0.385517 | 0.000000 |
| 116 | 14.05 | 8.616667 | 0.380147 | 0.000000 |
| 11 | 9.93 | 5.320000 | 0.348852 | 0.000000 |
| 15 | 4.14 | 2.276667 | 0.354805 | 0.000000 |
| 16 | 18.11 | 9.973333 | 0.355134 | 0.000000 |
| 27 | 9.67 | 6.580000 | 0.404923 | 0.000000 |
| 41 | 19.50 | 11.333333 | 0.367568 | 0.000000 |
| 54 | 8.57 | 4.596667 | 0.349114 | 0.000000 |
| 668 | 20.51 | 13.240000 | 0.392296 | 0.000000 |
| 67 | 13.01 | 8.406667 | 0.392529 | 0.000000 |
| 70 | 32.15 | 12.016667 | 0.272075 | 0.031746 |
| 202 | 31.00 | 12.000000 | 0.279070 | 0.036585 |
| 74 | 21.88 | 9.370000 | 0.299840 | 0.041667 |
| 71 | 27.73 | 13.103333 | 0.320898 | 0.043478 |
| 511 | 66.41 | 37.590000 | 0.361442 | 0.050000 |
| 227 | 53.80 | 37.533333 | 0.410949 | 0.052632 |
| 63 | 52.70 | 30.633333 | 0.367600 | 0.076923 |
| 426 | 116.87 | 64.380000 | 0.355200 | 0.105263 |

| | duree_stock_mois | stock_quantity | stock_status |
|-----|------------------|----------------|--------------|
| 2 | 1.200000e+07 | 0 | outofstock |
| 196 | 1.200000e+07 | 145 | instock |
| 622 | 1.200000e+07 | 0 | outofstock |
| 608 | 1.200000e+07 | 13 | instock |
| 536 | 1.200000e+07 | 0 | outofstock |
| 533 | 1.200000e+07 | 0 | outofstock |
| 490 | 1.200000e+07 | 0 | outofstock |
| 454 | 1.200000e+07 | 0 | outofstock |
| 393 | 1.200000e+07 | 0 | outofstock |
| 367 | 1.200000e+07 | 0 | outofstock |
| 248 | 1.200000e+07 | 0 | outofstock |
| 237 | 1.200000e+07 | 0 | outofstock |
| 212 | 1.200000e+07 | 0 | outofstock |
| 8 | 1.200000e+07 | 0 | outofstock |
| 201 | 1.200000e+07 | 97 | instock |
| 118 | 1.200000e+07 | 0 | outofstock |

| | | | |
|-----|--------------|-----|------------|
| 116 | 1.200000e+07 | 0 | outofstock |
| 11 | 1.200000e+07 | 0 | outofstock |
| 15 | 1.200000e+07 | 0 | outofstock |
| 16 | 1.200000e+07 | 0 | outofstock |
| 27 | 1.200000e+07 | 0 | outofstock |
| 41 | 1.200000e+07 | 0 | outofstock |
| 54 | 1.200000e+07 | 0 | outofstock |
| 668 | 1.200000e+07 | 0 | outofstock |
| 67 | 1.200000e+07 | 0 | outofstock |
| 70 | 3.779881e+02 | 125 | instock |
| 202 | 3.279910e+02 | 81 | instock |
| 74 | 2.879931e+02 | 71 | instock |
| 71 | 2.759937e+02 | 91 | instock |
| 511 | 2.399952e+02 | 19 | instock |
| 227 | 2.279957e+02 | 18 | instock |
| 63 | 1.559980e+02 | 12 | instock |
| 426 | 1.139989e+02 | 18 | instock |

Certains produits, présentent une combinaison critique : stock important, faible rotation et ancienneté élevée. Bien que leur marge soit correcte sur le papier, ils représentent une immobilisation logistique importante, sans retour immédiat. Ils doivent faire l'objet d'une surveillance spécifique, voire d'actions correctives ciblées.

4.2 Actions possibles :

Mettre en avant commercialement ces produits (homepage, mise en coffret, pack découverte)

Tester une promo douce ou bundle avec des vins qui tournent mieux

Réévaluer les visuels, le wording ou le référencement si invisibles sur le site

Prioriser leur inclusion dans une campagne newsletter ou vente privée

[]:

4.3 Création d'une alerte pour les produits dormant :

- `stock_quantity > 100`: On cherche les gros volumes de stock.
- `rotation_stock < 0.06`: Une rotation très faible.
- `duree_stock_mois > 36` : Le produit est en stock depuis plus de 3 ans. C'est un signe clair qu'il traîne.
- `total_sales < 10`: Moins de 10 ventes sur toute la période, c'est très peu et indique une faible demande.

```
[479]: erp_liaison_web['alerte_stock_dormant'] = (
    (erp_liaison_web['stock_quantity'] > 100) &
    (erp_liaison_web['rotation_stock'] < 0.06) &
    (erp_liaison_web['duree_stock_mois'] > 36) &
    (erp_liaison_web['total_sales'] < 10)
)
```

```
[481]: erp_liaison_web[ (
        (erp_liaison_web['stock_quantity'] > 100) &
        (erp_liaison_web['rotation_stock'] < 0.06) &
        (erp_liaison_web['duree_stock_mois'] > 36) &
        (erp_liaison_web['total_sales'] < 10)
    )]
```

```
[481]:
```

| | product_id | onsale_web | price | stock_quantity | stock_status | \ |
|-----|------------|------------|-------|----------------|--------------|---|
| 70 | 4142 | 1 | 53.0 | 125 | instock | |
| 75 | 4149 | 1 | 69.0 | 101 | instock | |
| 76 | 4150 | 1 | 59.0 | 123 | instock | |
| 194 | 4334 | 1 | 49.0 | 142 | instock | |
| 196 | 4337 | 1 | 83.0 | 145 | instock | |
| 197 | 4348 | 1 | 59.0 | 125 | instock | |
| 198 | 4350 | 1 | 79.5 | 145 | instock | |
| 203 | 4357 | 1 | 39.0 | 115 | instock | |
| 436 | 5024 | 1 | 45.0 | 103 | instock | |
| 437 | 5025 | 1 | 112.0 | 136 | instock | |
| 642 | 6126 | 1 | 135.0 | 138 | instock | |

| | purchase_price | id_web | sku | virtual | downloadable | ... | CA | \ |
|-----|----------------|--------|-------|---------|--------------|-----|-------|---|
| 70 | 32.15 | 11641 | 11641 | 0 | 0 | ... | 212.0 | |
| 75 | 40.25 | 7086 | 7086 | 0 | 0 | ... | 345.0 | |
| 76 | 35.45 | 1366 | 1366 | 0 | 0 | ... | 354.0 | |
| 194 | 30.01 | 7818 | 7818 | 0 | 0 | ... | 343.0 | |
| 196 | 48.90 | 4679 | 4679 | 0 | 0 | ... | 0.0 | |
| 197 | 34.76 | 12586 | 12586 | 0 | 0 | ... | 295.0 | |
| 198 | 47.30 | 12588 | 12588 | 0 | 0 | ... | 556.5 | |
| 203 | 22.30 | 9562 | 9562 | 0 | 0 | ... | 195.0 | |
| 436 | 27.04 | 11996 | 11996 | 0 | 0 | ... | 270.0 | |
| 437 | 68.60 | 13914 | 13914 | 0 | 0 | ... | 672.0 | |
| 642 | 80.33 | 14923 | 14923 | 0 | 0 | ... | 675.0 | |

| | z_score | price_ht | marge_unitaire | taux_de_marge | marge_totale | \ |
|-----|----------|------------|----------------|---------------|--------------|---|
| 70 | 0.749404 | 44.166667 | 12.016667 | 0.272075 | 48.066667 | |
| 75 | 1.329598 | 57.500000 | 17.250000 | 0.300000 | 86.250000 | |
| 76 | 0.966977 | 49.166667 | 13.716667 | 0.278983 | 82.300000 | |
| 194 | 0.604356 | 40.833333 | 10.823333 | 0.265061 | 75.763333 | |
| 196 | 1.837267 | 69.166667 | 20.266667 | 0.293012 | 0.000000 | |
| 197 | 0.966977 | 49.166667 | 14.406667 | 0.293017 | 72.033333 | |
| 198 | 1.710350 | 66.250000 | 18.950000 | 0.286038 | 132.650000 | |
| 203 | 0.241735 | 32.500000 | 10.200000 | 0.313846 | 51.000000 | |
| 436 | 0.459307 | 37.500000 | 10.460000 | 0.278933 | 62.760000 | |
| 437 | 2.888868 | 93.333333 | 24.733333 | 0.265000 | 148.400000 | |
| 642 | 3.722896 | 112.500000 | 32.170000 | 0.285956 | 160.850000 | |


```
rotation_stock duree_stock_mois alerte_produit_anomalie_critique \
```

| | | | |
|-----|----------|--------------|-------|
| 70 | 0.031746 | 3.779881e+02 | False |
| 75 | 0.049020 | 2.447950e+02 | False |
| 76 | 0.048387 | 2.479949e+02 | False |
| 194 | 0.048951 | 2.451378e+02 | False |
| 196 | 0.000000 | 1.200000e+07 | False |
| 197 | 0.039683 | 3.023924e+02 | False |
| 198 | 0.047945 | 2.502805e+02 | False |
| 203 | 0.043103 | 2.783935e+02 | False |
| 436 | 0.057692 | 2.079964e+02 | False |
| 437 | 0.043796 | 2.739937e+02 | True |
| 642 | 0.035971 | 3.335907e+02 | True |

| | alerte_stock_dormant |
|-----|----------------------|
| 70 | True |
| 75 | True |
| 76 | True |
| 194 | True |
| 196 | True |
| 197 | True |
| 198 | True |
| 203 | True |
| 436 | True |
| 437 | True |
| 642 | True |

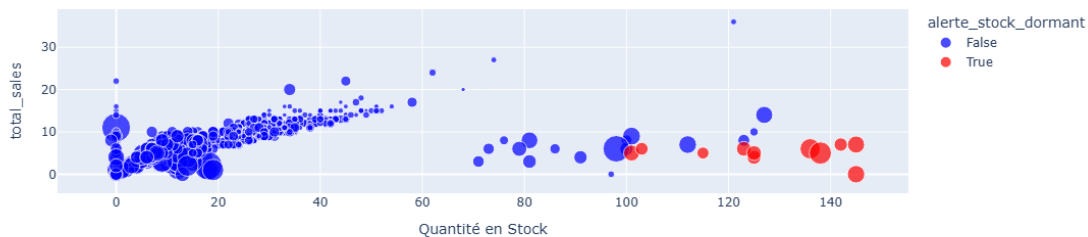
[11 rows x 41 columns]

[]:

```
[484]: fig = px.scatter(
    erp_liaison_web,
    x='stock_quantity',
    y='total_sales',
    color='alerte_stock_dormant',
    size='price',
    hover_name='post_title',
    title='<b>Analyse des Stocks : Identification des Produits Dormants</b>',
    labels={'stock_quantity': 'Quantité en Stock', 'duree_stock_mois': 'Durée_
↳ du Stock (Mois)'},
    color_discrete_map={True: 'red', False: 'blue'}
)

fig.update_layout(showlegend=True)
fig.show()
```

Analyse des Stocks : Identification des Produits Dormants



[]:

4.3.1 Corrélations

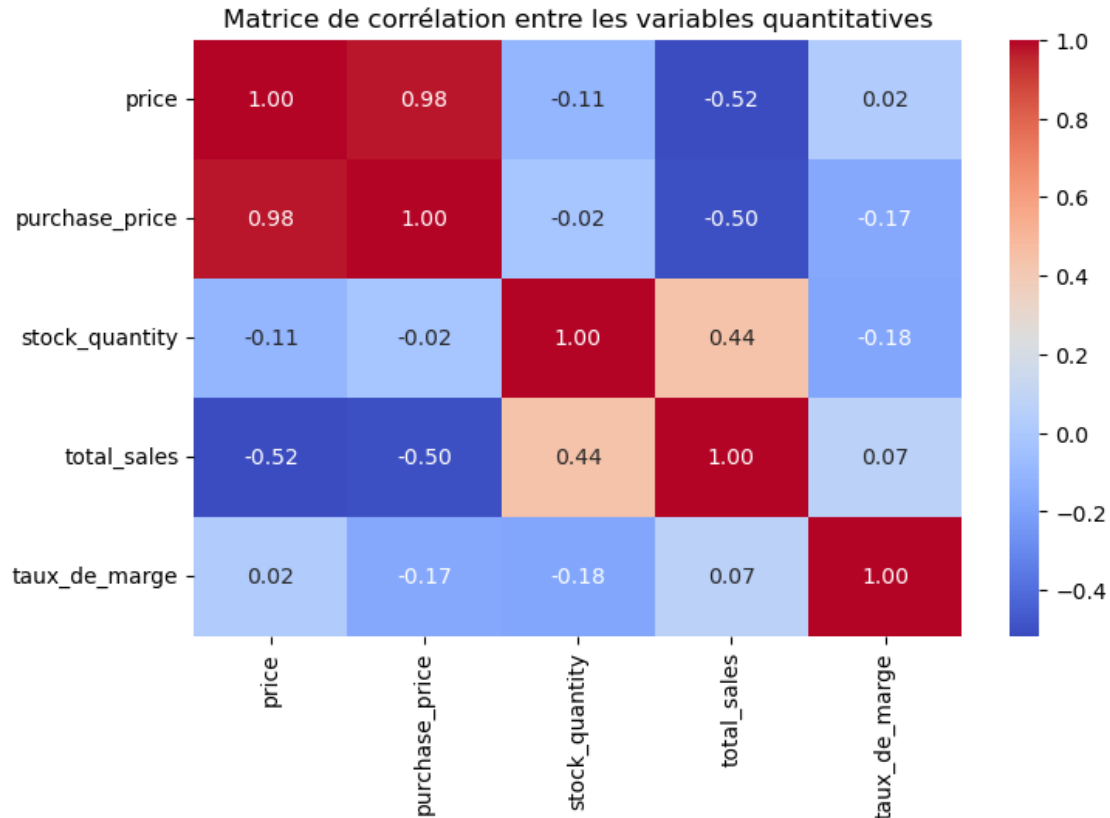
```
[488]: variables_quantitatives = ['price', 'purchase_price', 'stock_quantity',
    ↪ 'total_sales', 'taux_de_marge']
```

```
matrice_correlation = erp_liaison_web[variables_quantitatives].corr()
print(matrice_correlation)
```

| | price | purchase_price | stock_quantity | total_sales | \ |
|----------------|-----------|----------------|----------------|-------------|---|
| price | 1.000000 | 0.975958 | -0.106674 | -0.516258 | |
| purchase_price | 0.975958 | 1.000000 | -0.024053 | -0.503306 | |
| stock_quantity | -0.106674 | -0.024053 | 1.000000 | 0.438901 | |
| total_sales | -0.516258 | -0.503306 | 0.438901 | 1.000000 | |
| taux_de_marge | 0.017758 | -0.172237 | -0.182882 | 0.069645 | |

| | taux_de_marge |
|----------------|---------------|
| price | 0.017758 |
| purchase_price | -0.172237 |
| stock_quantity | -0.182882 |
| total_sales | 0.069645 |
| taux_de_marge | 1.000000 |

```
[490]: plt.figure(figsize=(8,5))
sns.heatmap(matrice_correlation, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Matrice de corrélation entre les variables quantitatives")
plt.show()
```



4.3.2 Analyse de la Matrice de Corrélation

La matrice de corrélation ci-dessus révèle les relations linéaires entre nos variables quantitatives clés. Les coefficients varient de -1 (corrélation négative parfaite) à +1 (corrélation positive parfaite), et 0 (absence de corrélation linéaire).

Voici les points saillants :

- **Prix d'Achat et Prix de Vente (price vs purchase_price) : Une très forte corrélation positive (0.98).** Cela est tout à fait logique : les produits achetés à un prix élevé sont, en règle générale, revendus à un prix élevé, reflétant la structure de coûts et de marge de l'entreprise.
- **Prix et Ventes (price/purchase_price vs total_sales) : Une corrélation négative modérée (-0.51 pour price, -0.50 pour purchase_price).** Cela indique que plus un produit est cher (à l'achat ou à la vente), moins il tend à se vendre en grande quantité. Ce comportement est typique sur le marché des vins et spiritueux, où les articles haut de gamme ont souvent un volume de vente plus faible que les produits d'entrée ou de milieu de gamme.
- **Quantité en Stock et Ventes Totales (stock_quantity vs total_sales) : Une corrélation positive modérée (0.44).** Cette relation suggère que les produits qui se vendent le plus (ou qui ont le potentiel de le faire) sont ceux pour lesquels l'entreprise maintient un niveau de stock plus élevé. Inversement, un stock faible pourrait être le reflet d'une faible

demande ou d'une gestion plus "juste-à-temps" pour les articles moins populaires.

- **Taux de Marge et autres variables (taux_de_marge vs toutes les autres) :** Le **taux de marge présente de très faibles corrélations** avec le prix de vente, le prix d'achat, la quantité en stock ou les ventes totales (tous les coefficients sont proches de 0). Ceci est une observation importante : la rentabilité relative (en pourcentage) d'un produit ne semble pas directement liée à son prix absolu, à son coût d'achat, ou à son volume de vente. L'entreprise pourrait avoir des stratégies de marge variées qui ne suivent pas un modèle linéaire simple, ou bien d'autres facteurs (comme la notoriété de la marque, les promotions spécifiques, etc.) influencent davantage la marge.

| Corrélation | Explication métier |
|---|---|
| price & purchase_price : forte taux_de_marge & autres : faible | Les produits chers à l'achat sont aussi chers à la vente Le taux de marge n'est pas fortement lié au volume ou au prix |
| total_sales & marge_totale : forte | Logique : plus on vend, plus on génère de marge globale |
| price & marge_totale : moyenne à faible | Un prix élevé n'implique pas nécessairement plus de marge globale |

[]: