

ANALYSE DU STOCK ET DES VENTES DU SITE BOTTLENECK



CONTENU



1. Analyses Exploratoires des Données
2. Fusion ou consolidations des données
3. Analyses univariées du prix
4. Analyses complémentaires (CA, quantités, stocks, taux de marge et corrélations)
5. Actions pour la suite
6. Point sur les compétences apprises



1. ANALYSES EXPLORATOIRES DES DONNÉES

01

DATASETS:

- Extraction ERP : Référence produit, prix, état du stock
- Extraction Site Web : SKU, quantités vendues, description des produits
- Table de Liaison : Lien entre la base de données Wordpress et l'ERP

02

CARACTÉRISTIQUES

- Erreurs Identifiées :
 - Stock_status n'est pas correctement remplie

stock_quantity	stock_status	purchase_price	stock_status_2	status_diff
3	outofstock	23.77	instock	False
0	instock	9.66	outofstock	False

- Valeurs négatives ou aberrantes:
 - Les stock_quantity et le price inférieurs à 0

```
Nombre d'article avec un prix non renseignés: 0
Le prix minimum de la colonne price: -20.0
Le prix maximum de la colonne price:225.0
Le prix inférieurs à 0:151 -20.0
69 -8.0
739 -9.1
Name: price, dtype: float64
```

```
La quantité minimum de la colonne stock_quantity:-10
La quantité maximum de la colonne stock_quantity:145
Les stocks inférieurs à 0:449 -10
673 -1
Name: stock_quantity, dtype: int64
```

- Incohérences dans les SKU:

'13127-1',

'bon-cadeau-25-euros',

Nan

- Valeur doublons dans les SKU avec la valeur de post_type sont different: 'product' et 'attachement'
- id_web sont pas unique: 90 duplicate et 91 null
- 91 product_id sans correspondances (id_web)

content_filtered	post_parent	guid	menu_order	post_type	post_mime_type
NaN	0.0	https://www.bottle-neck.fr/wp-content/uploads/...	0.0	attachment	image/jpeg
NaN	0.0	https://www.bottle-neck.fr/?post_type=product&...	0.0	product	NaN

1. ANALYSES EXPLORATOIRES DES DONNÉES

03

TRAITEMENT RÉALISÉS

- Corriger colonne stock_status
- Corriger colonne Price et stock: remplacées valeur négatif par la valeur absolue
- Gestion des doublons: Supprimer toutes les lignes ayant un 'sku' en double avec la valeur de post_type=='attachement'
- Uniformisation des formats : tous les sku sont des chaînes de caractères pour uniformité.
- Nettoyage des valeurs NaN: SKU
- Défini des colonnes à supprimer, effectuer l'opération

04

REMARQUE ÉVENTUELLES, PIÈGES OU DIFFICULTÉS RENCONTRÉES

- Analyser les raisons : Examiner le contexte de ces lignes pour comprendre pourquoi les prix sont négatifs. Est-ce dû à une erreur de saisie, à des retours de produits, à des ajustements comptables, ou autre?
- Corriger : Décider comment ces valeurs doivent être corrigées (par exemple, remplacées par zéro ou la valeur absolue)

2.FUSION OU CONSOLIDATIONS DES DONNÉES

01

CHOIX DES ATTRIBUTS

- Extraction ERP :
 - product_id: Identifiant unique du produit dans l'ERP.
- Extraction Site Web :
 - sku: Stock Keeping Unit, identifiant unique du produit sur le site Web.
- Table de Liaison :
 - product_id: Identifiant unique du produit dans l'ERP.
 - id_web: Identifiant unique du produit sur le site Web.

02

CLÉ UTILISÉES

- Fusion ERP et table de liaison:
 - Clé de jointure: 'product_id
- Fusion résultat précédent et site web:
 - Clé de jointure: id_web et sku

2.FUSION OU CONSOLIDATIONS DES DONNÉES

03

VIGILANCE PARTICULIÈRES AU COURS DU TRAITEMENT

- Identification Incohérences et Correspondances entre les Clés:
 - Vérifier que les clé correspondent bien entre les différents tables.
 - Identifier et traiter les valeurs nulles ou manquantes avant la fusion
- Formats et types de données:
 - S'assurer que les clés utilisées pour la jointure sont du même type
- Doublons et unicité des identifiants :
 - Vérifier l'unicité des identifiants dans chaque table avant la fusion
 - Supprimer les doublons potentiels pour éviter des erreurs de fusion

04

DIFFICULTÉS OU PIÈGES RENCONTRÉS

Jonction du fichier df_erp et df_liaison

- Gestion des Valeurs Nulles :
 - La présence de valeurs nulles dans les clés de jointure peut poser des problèmes.
- Correspondances Incomplètes :
 - Certaines références (product_id, id_web) peuvent ne pas avoir de correspondance dans toutes les tables, ce qui peut mener à des pertes de données lors des fusions.

	purchase_price	id_web
19	37.88	nan
49	33.79	nan
50	25.25	nan
119	7.36	nan
131	33.01	nan
..
817	31.20	nan
818	15.54	nan
819	16.02	nan
820	23.48	nan
821	24.18	nan

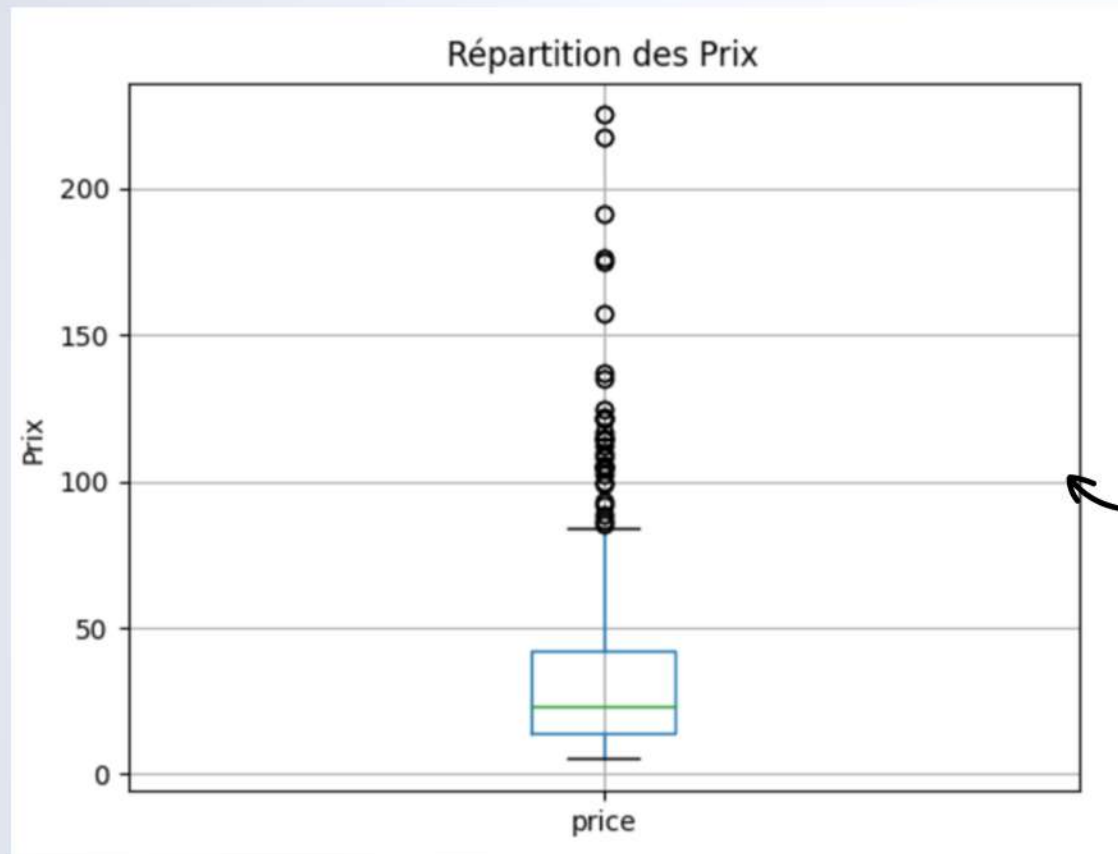
[91 rows x 7 columns]

3. ANALYSES UNIVARIÉES DU PRIX

01

EXPLORATION PAR LA VISUALISATION DE DONNÉES

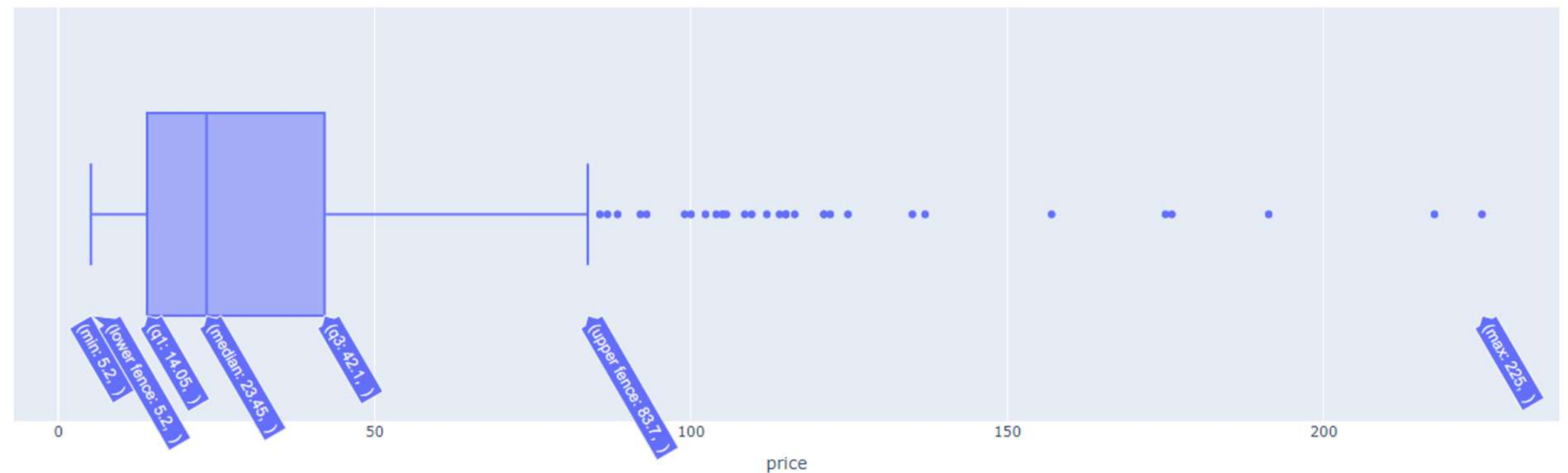
- Création d'une boîte à moustache grâce à:



matplotlib.pyplot

plotly express

La répartition des prix



3. ANALYSES UNIVARIÉES DU PRIX

02

EXPLORATION PAR L'UTILISATION DE MÉTHODES STATISTIQUES

Identification les outlier par le Z- index

```
#Calculer la moyenne du prix
mean=round(np.mean(df_final['price']),2)
print('la moyenne du prix: {}'.format(mean))
#Calculer l'écart-type du prix
ecart_type= round(np.std(df_final['price']),2)
print('l\'écart-type du prix: {}'.format(ecart_type))
#Calculer le Z-score
df_final['Z-score']= round((df_final['price']- mean)/ecart_type,3)
df_final
```

la moyenne du prix: 32.33
l'écart-type du prix: 27.58

Le seuil prix dont z-score est supérieur à 3: 116.4

Des outliers potentiellement intéressants à analyser plus en détail

Outlier basé sur le prix:

	price	Z-score
140	175.0	5.173
141	175.0	5.173
216	225.0	6.986
217	225.0	6.986
242	116.4	3.048
243	116.4	3.048
436	122.0	3.251
437	122.0	3.251
546	124.8	3.353
547	124.8	3.353
704	157.0	4.520
705	157.0	4.520
744	137.0	3.795
745	137.0	3.795
904	176.0	5.209
905	176.0	5.209
974	191.3	5.764
975	191.3	5.764
1104	217.5	6.714
1105	217.5	6.714
1142	135.0	3.723
1143	135.0	3.723
1168	121.0	3.215
1169	121.0	3.215
1406	121.0	3.215
1407	121.0	3.215

3. ANALYSES UNIVARIÉES DU PRIX



02

EXPLORATION PAR L'UTILISATION DE MÉTHODES STATISTIQUES

Identification par l'interval interquartile

```
df_final['price'].describe()

count    1428.000000
mean      32.333683
std       27.586661
min        5.200000
25%       14.050000
50%       23.450000
75%       42.100000
max      225.000000
Name: price, dtype: float64
```

```
#Calcul de l'IQR(interquartile range): est la différence e
Q1= 14.05
Q3= 42.10
IQR= Q3-Q1
#Seuils pour les Outliers:
seuil_supérieur= Q3+ 1.5*IQR
seuil_inferieur= Q1- 1.5*IQR
print(f"Seuil supérieur pour outlier: {seuil_supérieur}")
print(f"Seuil inférieur pour outlier: {seuil_inferieur}")

Seuil supérieur pour outlier: 84.17500000000001
Seuil inférieur pour outlier: -28.025000000000002
```

Nombre d'articles du catalogue outliers:31
La proportion de l'ensemble du catalogue outliers: 4.34%

#Comparer avec des produits similaires sur d'autres marchés ou dans d'autres bases de données

```
product_type
Champagne    139.000000
Cognac       136.250000
Vin          119.222222
Whisky       109.666667
Name: price, dtype: float64
```

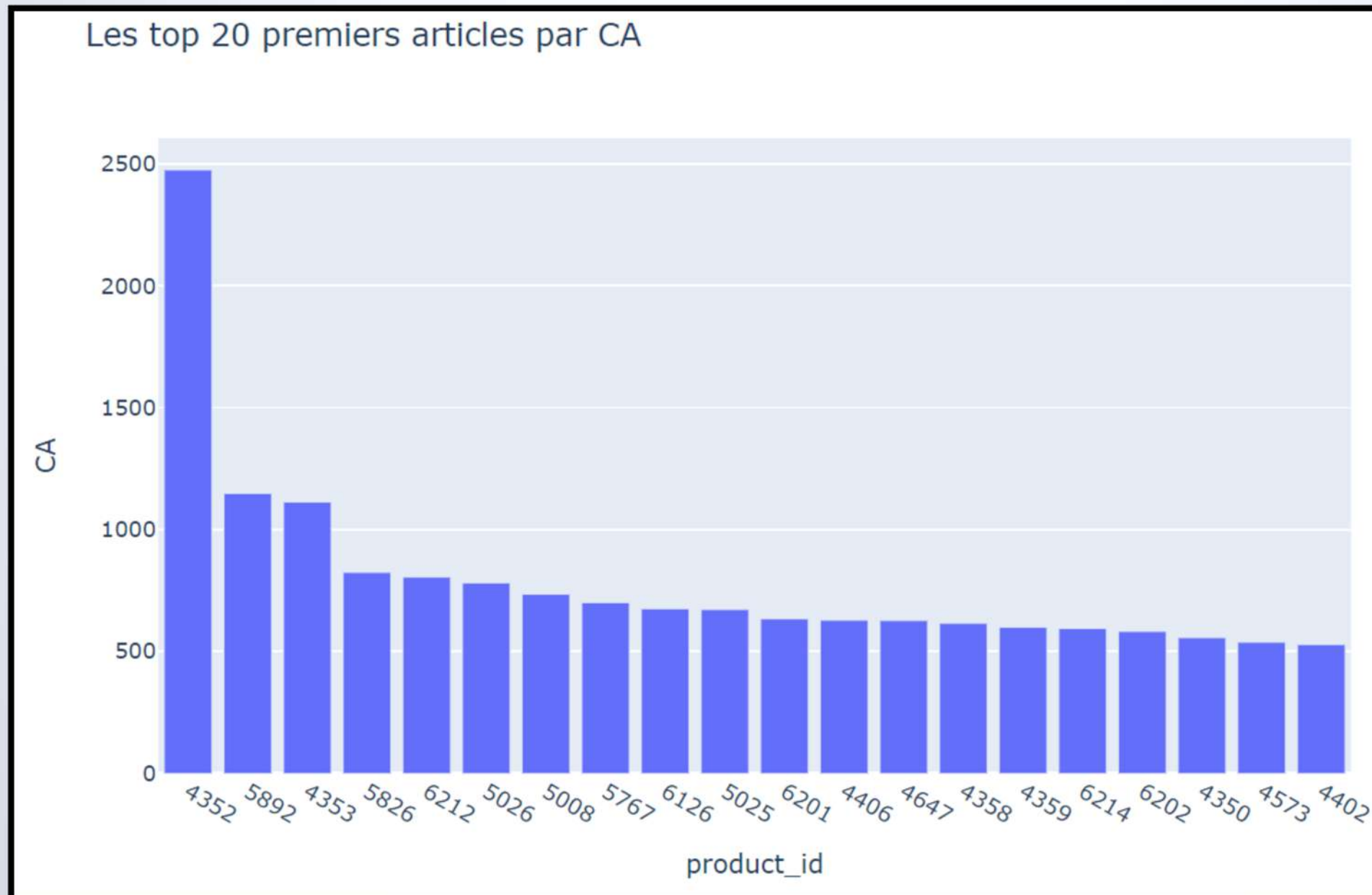
Les prix sur votre site sont plus élevés ce qui suggère une concentration sur des produits haut de gamme et de qualité supérieure.

4.ANALYSES DES CHIFFRE AFFAIRE ET VENTES EN QUANTITÉS

Méthode d'analyse statistique

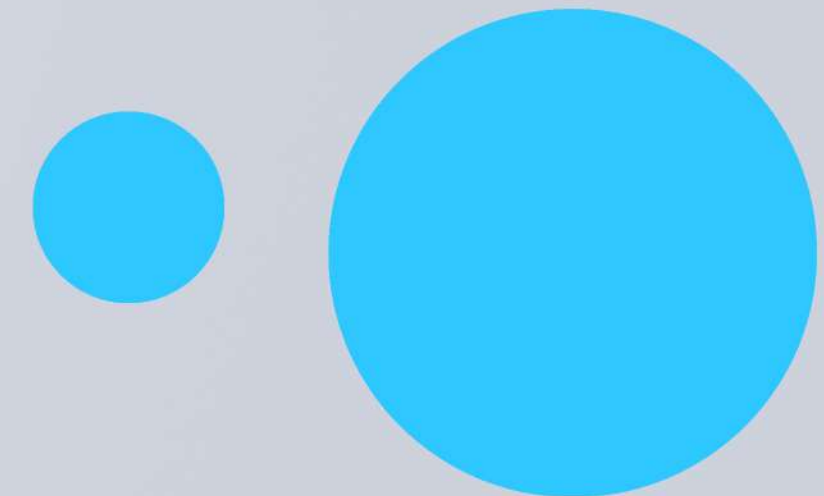
Analyse de Pareto (Règle du 20/80):

- Dans de nombreux cas, environ 80 % des effets proviennent de 20 % des causes.
- Aide à identifier les articles les plus significatifs en termes de contribution au CA et ventes.

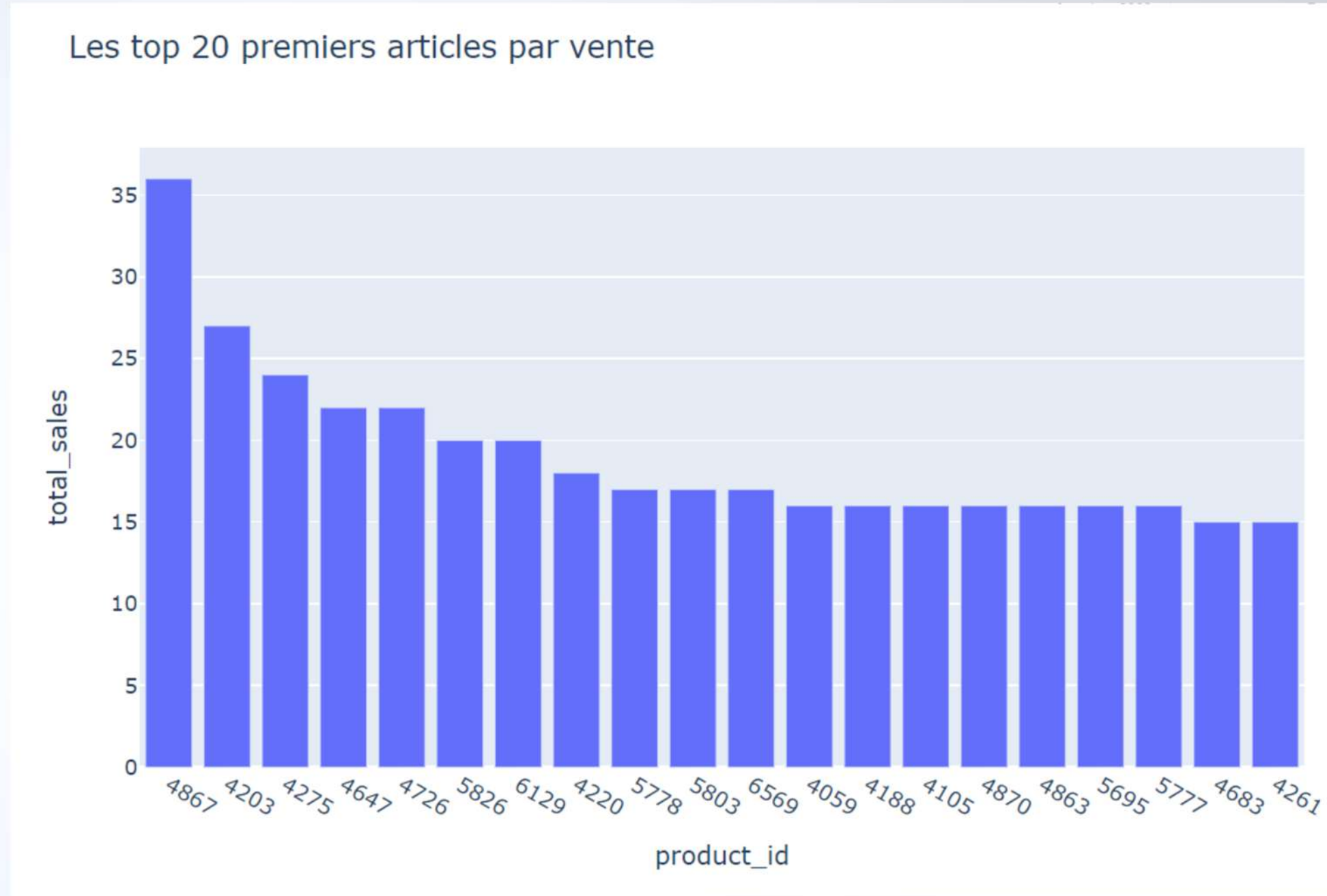


Le chiffre d'affaire du site web :143680.1

Le nombre d'articles représentant 80% du CA: 434
La proportion des articles représentant 80% du CA:60.78 %



4.ANALYSES DES CHIFFRE AFFAIRE ET VENTES EN QUANTITÉS



Le nombre d'articles représentant 80% des ventes en quantité: 432
La proportion d'articles dans le catalogue entier du site web: 60.5%

4.ANALYSES DES STOCKS

Graphique en barre du flop 20 des produits qui ont le plus de mois de stock



La valorisation de notre stock en euro est de 494682.4 €

La valorisation du nombre de produit en stock: 16741

4.ANALYSES DE TAUX DE MARGE BRUT

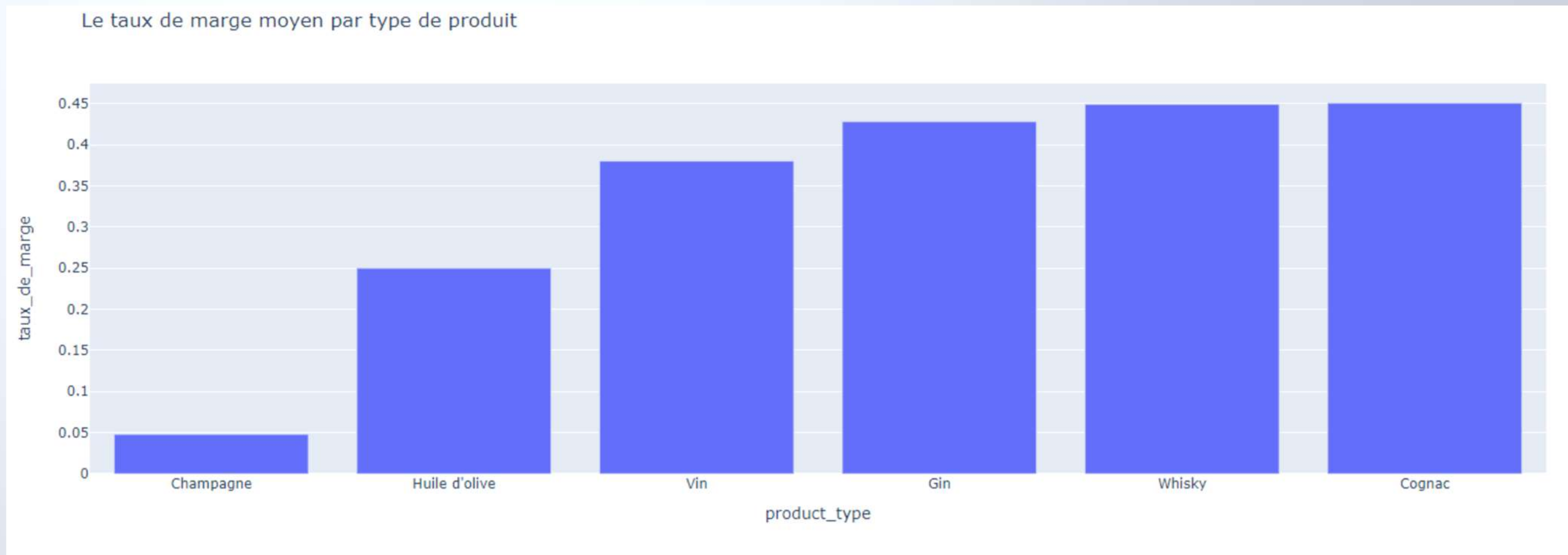
#Calcul de la marge brute

duct_type	comment_count	product_id	onsale_web	price	...	purchase_price	id_web	Z-score	CA	COGS	rotation_stock	Valorisation_stock_euros	prixHT	marge_brut	taux_de_marge
champagne	0.0	4355	1	12.65	...	77.48	12589	-0.714	0.0	7515.56	320.536995	1227.05	10.541667	-66.93833	-6.349881

Le taux de marge minimum: 0.23
Le prix de la taux de marge minimum:[13.1]

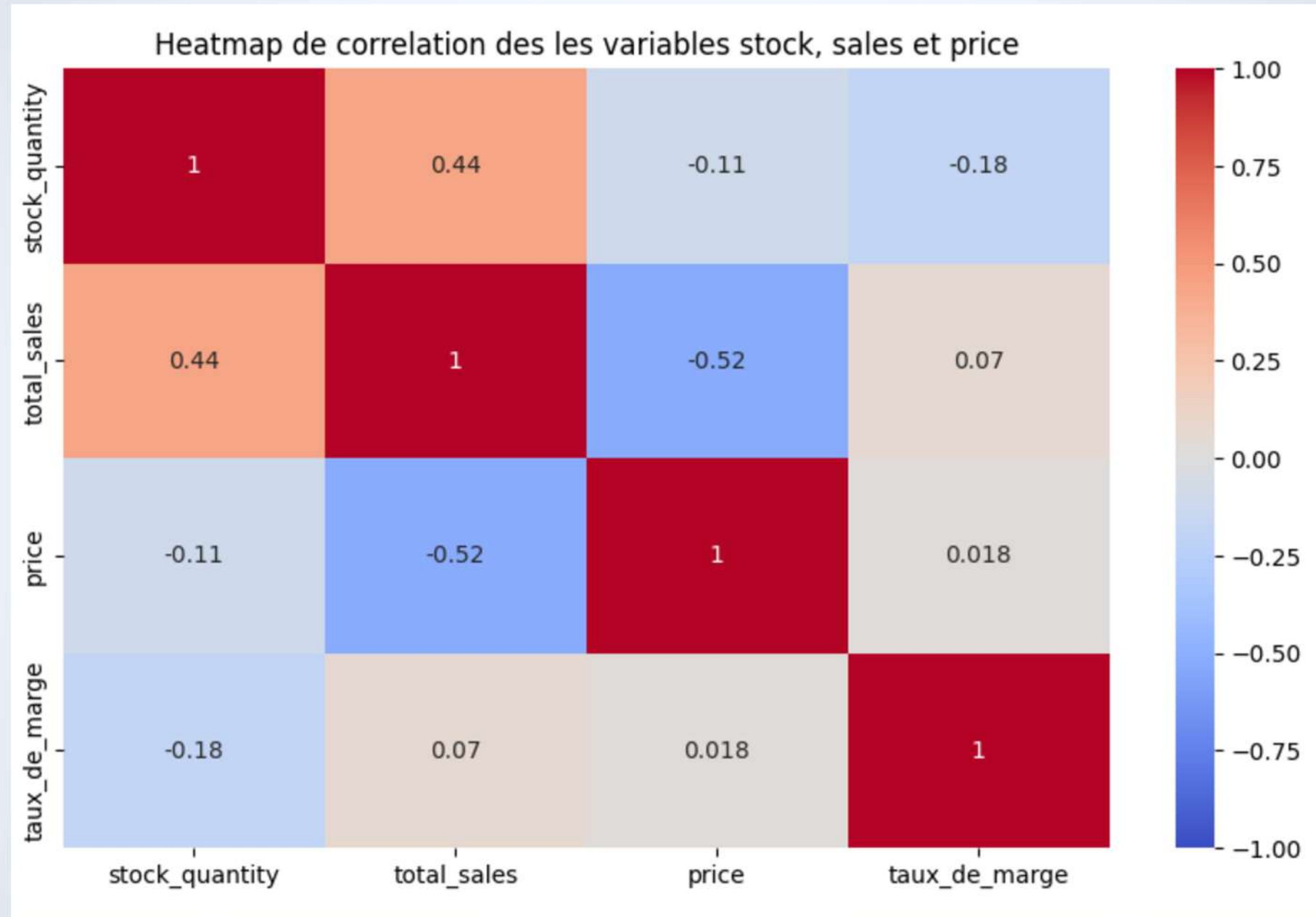
Le taux de marge maximum: 0.48
Le prix de la taux de marge maximum:[62.5]

#Comparez les marges brutes entre différents produits



4. ANALYSE DES CORRELATIONS ENTRE LES VARIABLES STOCK, SALES, PRICE ET TAUX DE MARGE

#Graphique en barre du flop 20 des produits qui ont le plus de mois de stock



- La relation la plus notable est entre total_sales et price -> des prix plus bas sont associés à des ventes plus élevées.
- Les autres corrélations sont faibles

#Limites éventuelles de l'analyse : identifier des relations potentielles entre différentes variables, mais il est important de noter que la corrélation ne signifie pas causalité.

5.ACTIONS POUR LA SUITE



AMÉLIORATION DE LA QUALITÉ DES DONNÉES

- Standardiser les formats de données
- Mettre à jour les tables de liaison pour les nouveaux produits



OPTIMISATION DES STOCK

- Revoir les stratégies d'approvisionnement et de commande pour éviter l'accumulation future de produits à faible rotation
- Surveiller continuellement les niveaux de stock en fonction des tendances de vente et de la demande réelle.s



STRATÉGIES DE VENTE ET MARKETING

- Focus sur les produits à haute marge pour maximiser les profits.
- Reconsidérer les produits à faible marge en termes de stratégie de prix, de promotion, ou de réduction des coûts.

6.POINT SUR LES COMPÉTENCES APPRISSES

POINT POSITIFS DU NETTOYAGE

- Clarté et organisation des données
- Utilisation de pandas pour le traitement des données
- Résolution des problème courants

DIFFICULTÉS RENCONTRÉES

- Gestion des incohérences entre sources
- Traitement des valeurs manquantes
- Complexité des données

TÂCHES BESOINS D'ENTRAÎNEMENT?

- Gestion des incohérences de données
- Technique avancées de nettoyage
- Optimisation des processus de traitement des données

MERCI POUR VOTRE ATTENTION

