

Optimisez la gestion des données d'une boutique

Par OpenClassRooms

Contexte et spécification des données



Marchand de vin prestigieux et autres spiritueux

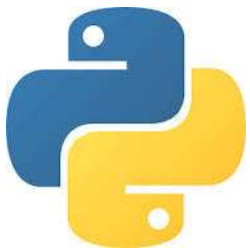
Optimisez la gestion des données d'une boutique

Reprendre la base de données, les outils étant artisanaux

- Identifier les erreurs dans les données
- Faire une liaison entre la base de données ERP et Web
- Calculer le chiffre d'affaires par produit/total
- Analyser l'état, la rotation des stocks et les taux de marge
- Faire une corrélation entre certaines données quantitatives

Méthodologie de l'analyse

Outils utilisés



Langage de programmation, conçu pour les analyses de données



Jupyter est un environnement de développement interactif basé sur le web pour les notebooks, le code et les données.

Ressources



ERP.xlsx



Web.xlsx



Liaison.xlsx

Nettoyages des données

Analyse exploratoire du fichier ERP

- Erreur de saisie colonne 'stock_status'

product_id	onsale_web	price	stock_quantity	stock_status	purchase_price
4039	1	46	3	outofstock	23,77
4885	1	18,7	0	instock	9,66

- prix négatives dans la colonne 'price'

product_id	onsale_web	price	stock_quantity	stock_status	purchase_price
4233	0	-20	0	outofstock	10,33
6594	0	-9,1	19	instock	4,61
5017	0	-8	0	outofstock	4,34

- stock négative dans la colonne 'stock_quantity'

product_id	onsale_web	price	stock_quantity	stock_status	purchase_price
4973	0	10	-10	outofstock	4,96
5700	1	44,5	-1	outofstock	22,3

Nettoyages des données

Analyse exploratoire du fichier ERP

- Filtre sur la valeur 0 dans la colonne 'onsale_web', il n'y a pas de correspondance avec les autres tables, pas de vente.

id_web	product_id
	6866
	6869
	6875
15735	6884
14897	6886
15736	6887
	6898
	6899



product_id	onsale_web
6866	0
6869	0
6875	0
6884	1
6886	1
6887	1
6898	0
6899	0

- Prix d'achat supérieur au prix de vente

product_id	onsale_web	price	stock_quantity	stock_status	purchase_price
4233	0	-20	0	outofstock	10,33
4355	1	12,65	97	instock	77,48
4864	0	8,3	0	outofstock	9,99
6594	0	-9,1	19	instock	4,61
6324	0	92	18	instock	99
5017	0	-8	0	outofstock	4,34
7196	0	31	55	instock	31,2

Vendre un produit moins cher que ce qu'on l'a acheté est appelé la vente à perte. Cette pratique est interdite.

Nettoyages des données

Analyse exploratoire du fichier Web

- Garder uniquement les produits pour enlever les doublons

sku	total_sales	product_type	post_title	post_type	post_mime_type
38	10	Vin	Emile Boeckel Crémant Brut Blanc de Blancs	attachment	image/jpeg
38	10	Vin	Emile Boeckel Crémant Brut Blanc de Blancs	product	
41	6	Vin	Marcel Windholtz Eau de Vie de Marc de Gewurztraminer	product	
41	6	Vin	Marcel Windholtz Eau de Vie de Marc de Gewurztraminer	attachment	image/jpeg

- Valeurs sans code articles avec des ventes négatives.

sku	total_sales	product_type	post_title	post_type
	-56	Vin	Pierre Jean Villa Condrieu Jardin Suspendu 2018	product
	-17	Vin	Pierre Jean Villa Côte Rôtie Fongéant 2017	product

- Remplacer la valeur manquante dans la colonne 'product_type' par 'Autre'

sku	total_sales	product_type	post_title	post_type
bon-cadeau-25-euros	7	Autre	Bon cadeau de 25€	attachment
bon-cadeau-25-euros	7		Bon cadeau de 25€	product

- ID doit être un nombre entier pour une meilleure gestion des articles

sku	total_sales	product_type	post_title	post_type
bon-cadeau-25-euros	7		Bon cadeau de 25€	product
13127-1	4	Vin	Clos du Mont-Olivet Châteauneuf-du-Pape 2007	product

Nettoyages des données

Analyse exploratoire du fichier Liaison

- Valeurs sans correspondances

id_web	product_id
	4055
	4090
	4092
	4195
	4209
	4233
	4278
	4279
	4565
	4577
	4578
	4594
	4599
	4659
	4692
	4693
	4697
	4698

Consolidations des données

Jonction du fichier df_erp et df_liaison

```
#Fusion des fichiers df_erp et df_liaison
df_merge = pd.merge(df_liaison, df_erp_onsale, on='product_id', how='inner')
df_merge
```

Clé utilisé :
Product_ID

	id_web	product_id	index	onsale_web	price	stock_quantity	stock_status	purchase_price
0	15298	3847	403	1	24.2	16	instock	12.88

Jonction du fichier df_merge et df_web

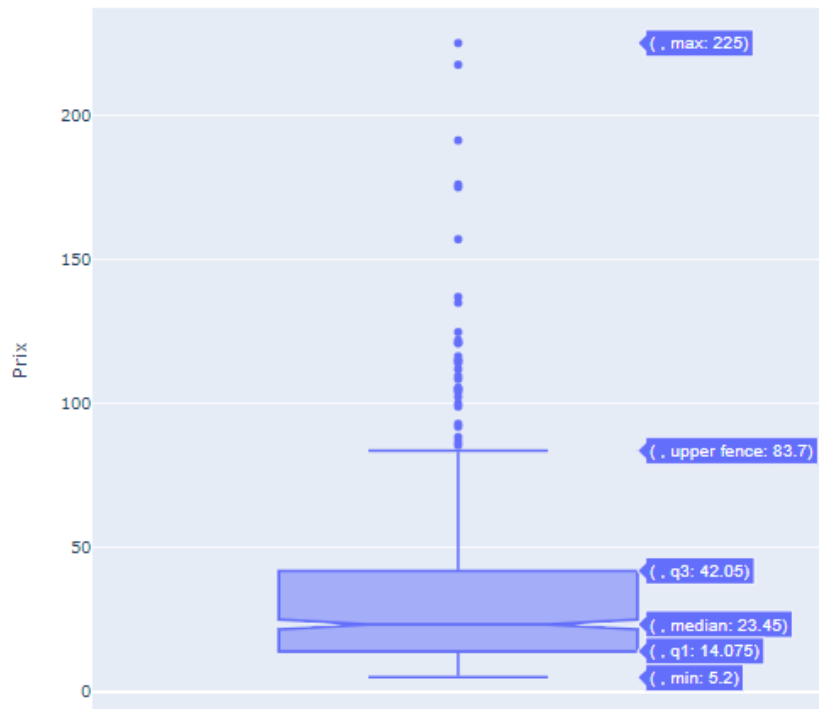
```
#Renommer la colonne 'sku' par 'id_web'
df_web = df_web.rename(columns={'sku':'id_web'})
#Fusionnez les datasets df_merge et df_web
df_merge2 = pd.merge(df_merge, df_web, on='id_web', how='left')
```

Clé utilisé : Id_web

	id_web	product_id	index	onsale_web	price	stock_quantity	stock_status	purchase_price	total_sales	tax_status	post_date	product_type
0	15298	3847	403	1	24.2	16	instock	12.88	6.0	taxable	2018-02-08	Vin

Analyse univarié des prix

Boite à moustache de la répartition des prix



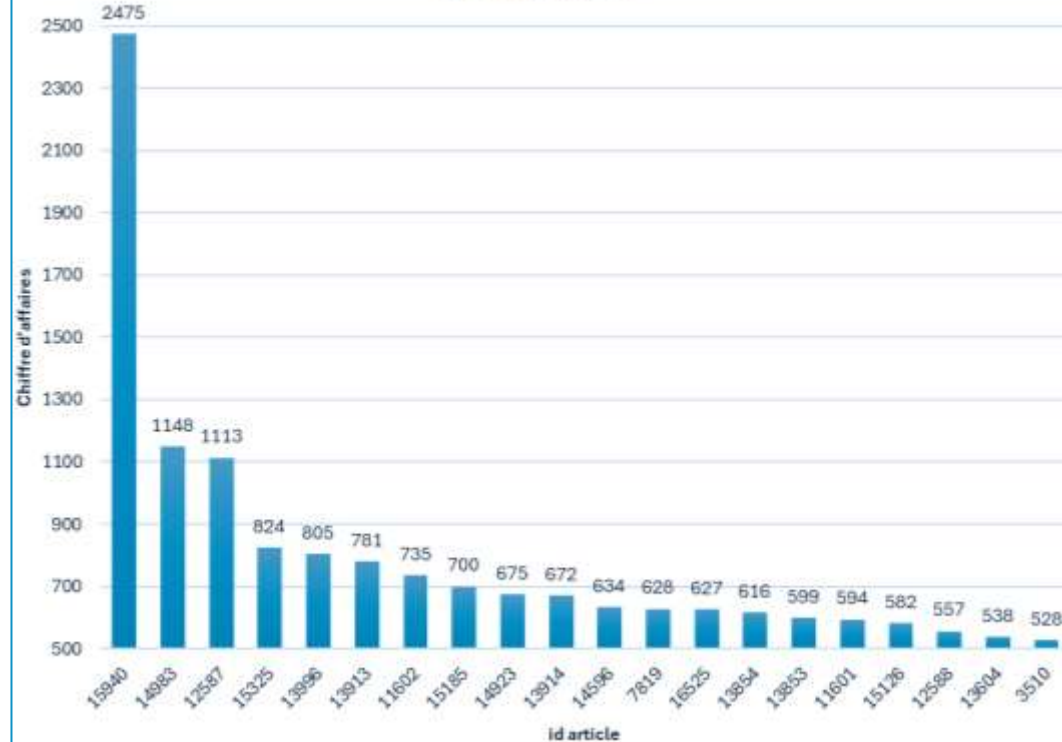
- La médiane est de 23,45€.
- 50% des prix sont compris entre 14 et 42€.
- Les outliers sont supérieurs à 83,7€.
- Il y a 31 articles dans le catalogue "outliers" ce qui représentent 4.35 % du catalogue total.

	post_title	price
63	Zind-Humbrecht Riesling Grand Cru Rangen De Th...	100.0
65	Zind-Humbrecht Pinot Gris Grand Cru Rangen De ...	88.4
198	Champagne Egly-Ouriet Grand Cru Millésimé 2008	225.0
204	Champagne Larmandier-Bernier Grand Cru Les Che...	85.6
217	Cognac Frapin VIP XO	176.0

D'après le tableau, la majorité des prix élevés qui se trouvent dans les outliers sont des grands crus dont certaines bouteilles sont réputées pour leurs qualités exceptionnelles. Les tarifs sont donc justifiés.

Analyse univarié du chiffre d'affaires

ca_par_article

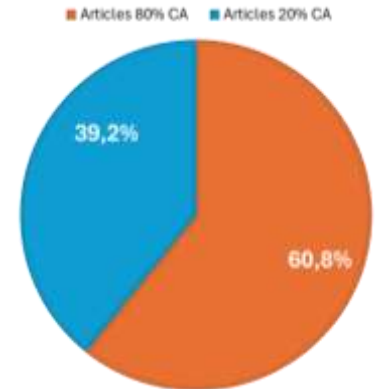
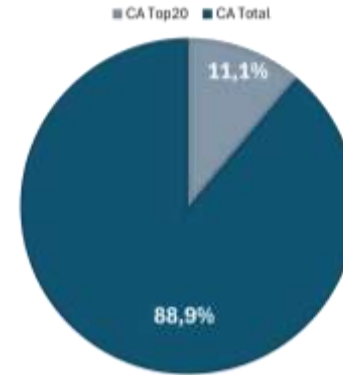


CA total = 143 242 €

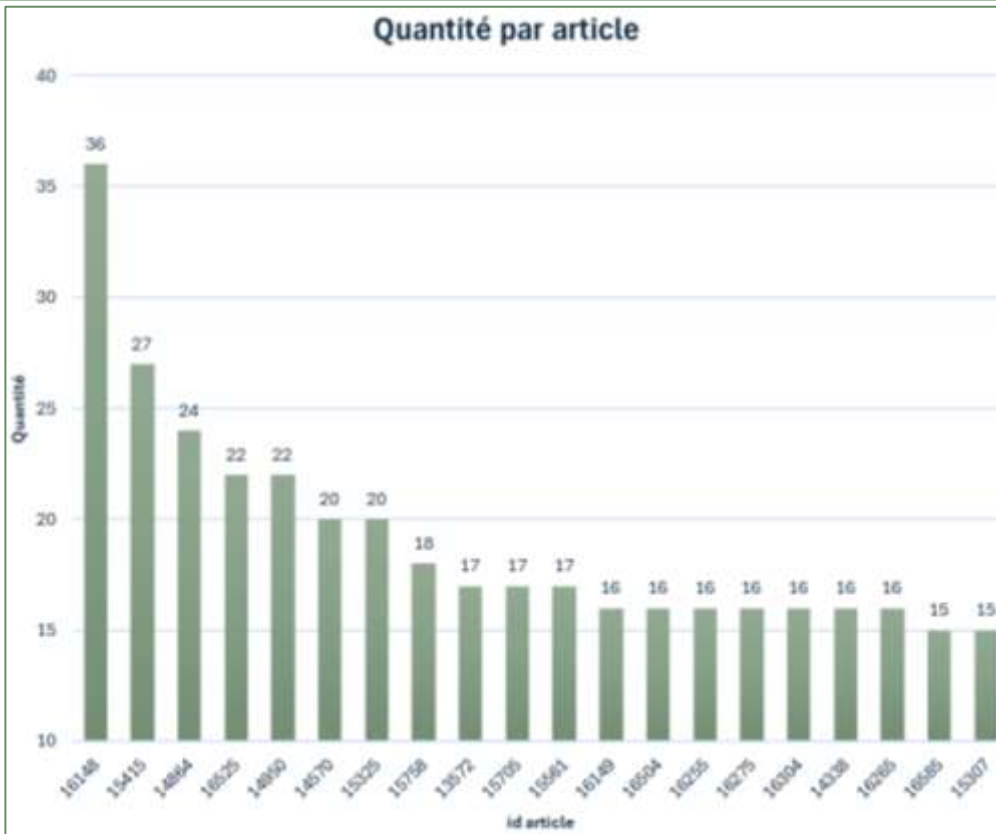
Nombre d'articles
totale : 712

CA du top 20 produits :
15 900 €

Nombre d'articles
représentant 80% du
CA : 433



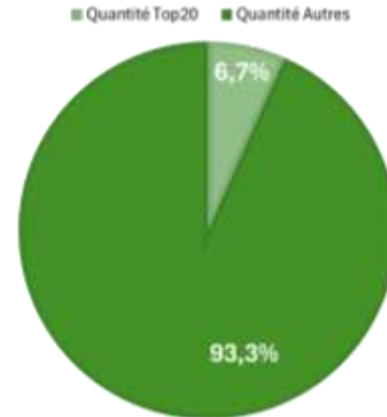
Analyse univarié des ventes en quantités



Quantité totale : 5 729 articles

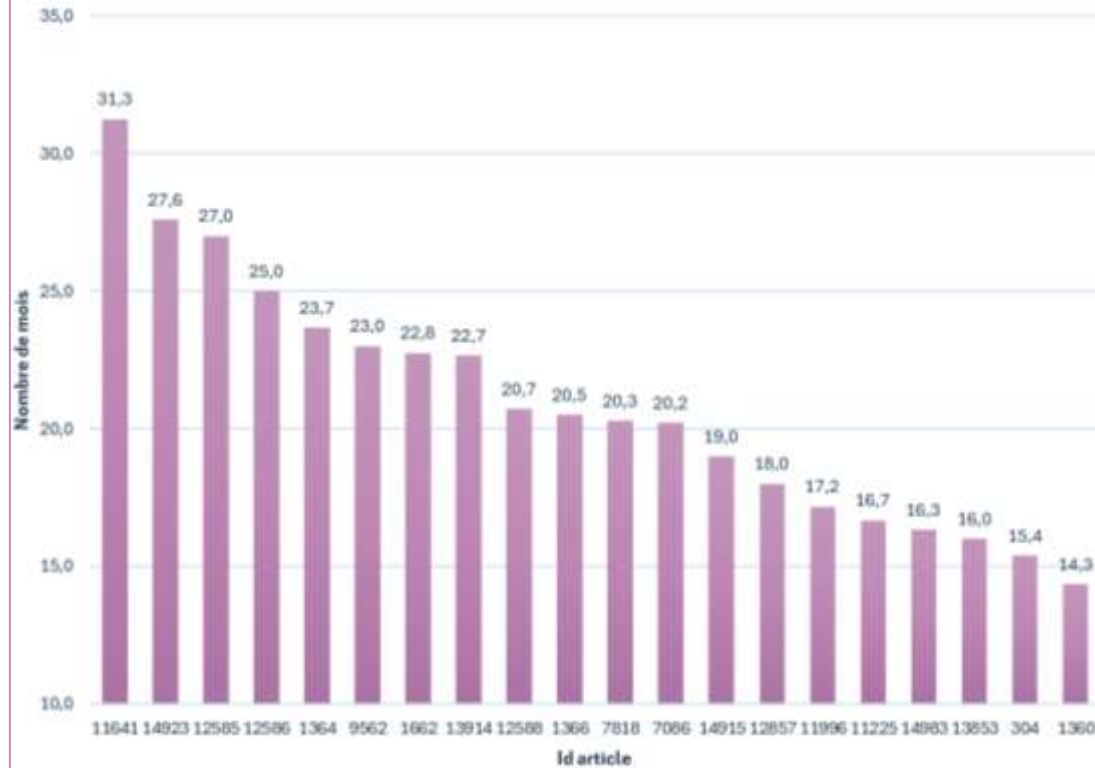
Quantité du top 20 :
382 articles

Nombre d'articles
représentant 80% des
ventes en quantité: 431



Rotation des stocks

Produits avec le plus de mois de stock



Rotation des
stocks moyens

2,98 = 3 mois

Valorisation des
stocks

493 219 €

Quantité en
stock

16 610 articles

Analyse du taux de marge

En HT

CALCUL DU
TAUX DE MARGE

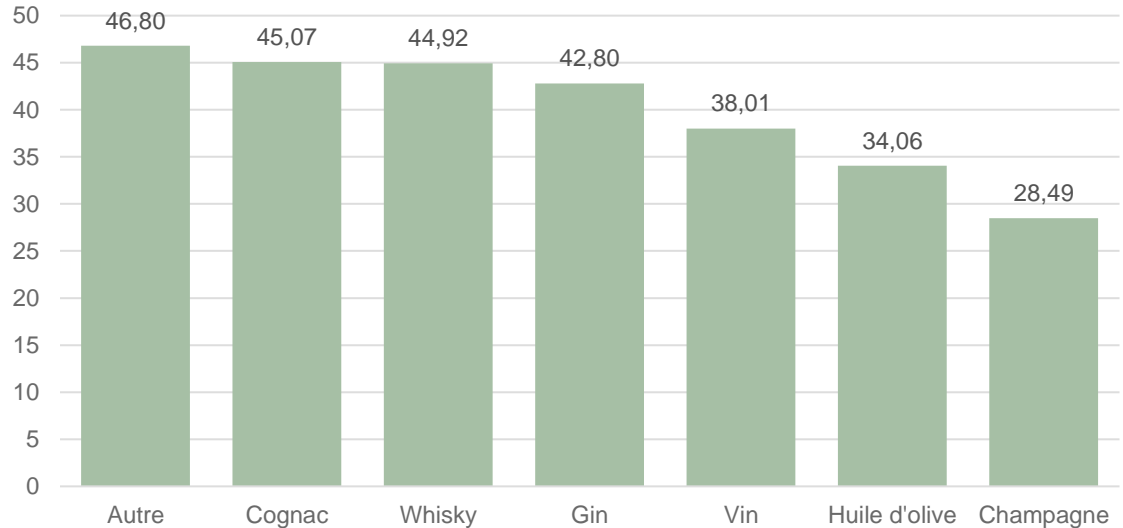


MARGE BRUTE /
COÛTS D'ACHAT X 100

Tableau pour le calcul des prix en HT

Produits	Taux de TVA
Huile d'olive	5.5%
Champagne	20%
Vin	20%
Cognac	20%
Whisky	20%
Gin	20%
Autres	0%

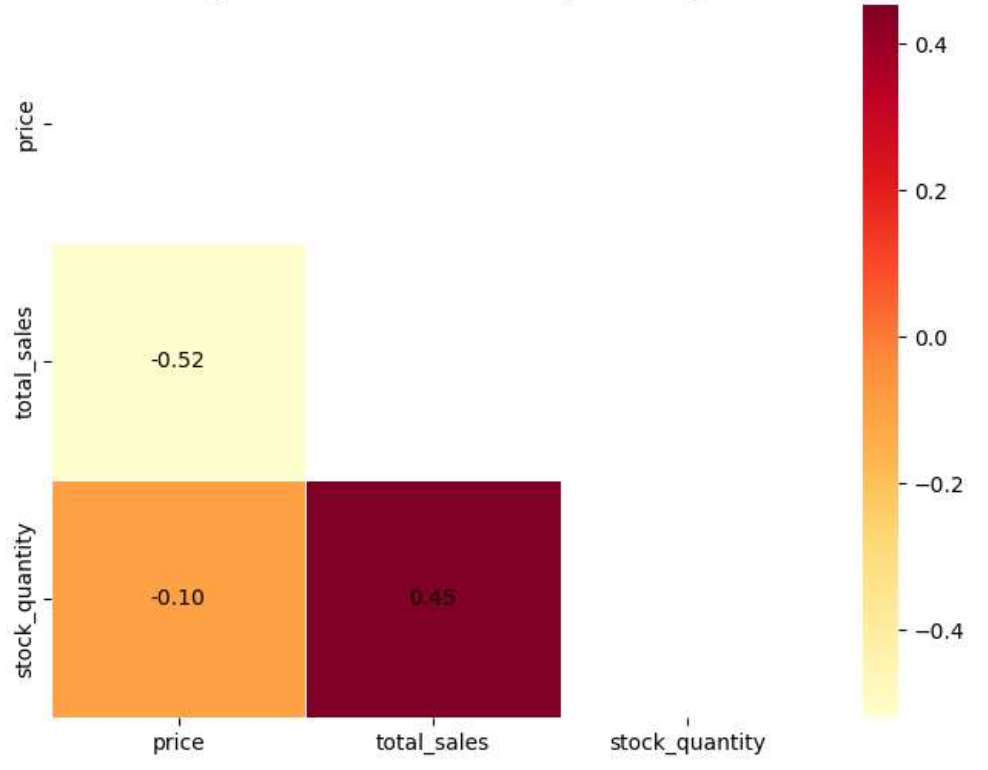
Tableau pour le calcul des prix en HT



- ❖ Le taux de marge moyen pour les articles est de 37.88%
- ❖ Les meilleurs taux de marge sont pour les boissons spiritueuses (Cognac, Whisky, Gin)

Analyse des corrélations

Heatmap de corrélation entre stock, sales et price



➤ Corrélation entre 'sales' et 'Price'

Corrélation négative modérée

Lorsque le prix des produits augmente, les ventes totales ont tendance à diminuer

- 0.52

➤ Corrélation entre 'Stock' et 'Price'

Corrélation négative faible

Légère tendance à ce que les produits avec des prix plus élevés aient une quantité en stock légèrement plus faible

- 0.10

➤ Corrélation entre 'sales' et 'Stock'

corrélation positive modérée

Lorsque la quantité en stock augmente, les ventes totales ont tendance à augmenter également

0.45