



# Gestion des données prix

Rapport sur l'étude des prix pratiqués

09/10/2023

# Fichier ERP

	product_id	onsale_web	price	stock_quantity	stock_status
0	3847	1	24.2	0	outofstock
1	3849	1	34.3	0	outofstock
2	3850	1	20.8	0	outofstock
3	4032	1	14.1	0	outofstock
4	4039	1	46.0	0	outofstock
5	4040	1	34.3	0	outofstock
6	4041	1	32.7	15	instock
7	4042	1	31.2	34	instock
8	4043	1	60.0	12	instock
9	4045	1	42.6	66	instock
10	4046	1	80.0	0	outofstock

825 lignes

2 erreurs sur des prix qui sont négatifs

= > on les remplace par la médiane pour pouvoir travailler les données

2 erreurs sur des stock qui sont négatifs

= > on les remplace par 0 car en adéquation avec outofstock

Correction des erreurs de corrélation entre quantité en stock et statut du stock

# Fichier web

	sku	virtual	downloadable	rating_count	average_rating	total_sales	tax_status	tax_class	post_author	post_date	...	post_name	post_modified
751	14253.0	0	0	0	0.0	0.0	taxable	NaN	2.0	2018-02-20 15:29:32	...	domaine-huet-vouvray-haut-lieu-sec-2016	2020-07-22 09:55:01
604	16154.0	0	0	0	0.0	4.0	taxable	NaN	2.0	2018-02-28 13:21:40	...	chateau-de-cazeneuve-pic-saint-loup-roc-des-ma...	2020-08-22 10:25:02
1226	15161.0	0	0	0	0.0	0.0	taxable	NaN	2.0	2019-07-30 14:59:19	...	le-pas-de-lescalette-terrasses-du-larzac-pas-d...	2020-02-10 09:30:09
1225	16119.0	0	0	0	0.0	1.0	taxable	NaN	2.0	2018-04-17 21:44:24	...	albert-mann-riesling-grand-cru-schlossberg-2018	2020-07-28 09:30:17
1223	3507.0	0	0	0	0.0	2.0	taxable	NaN	2.0	2018-03-22 11:32:55	...	cognac-frapin-fontpinot-xo	2020-08-12 09:30:16

1513 lignes au départ

4 erreurs sur le sku :

= > 'bon cadeau 25 euros'

= > '13127-1'

= > présence de caractères considérés comme lettres dans 2 sku

= > 2 sku absents mais avec des lignes complètes

35 lignes sans aucunes informations

714 doublons sku

= > Suppression des lignes les moins biens référencées

716 lignes utilisables après traitement

# Fichier liaison

	product_id	id_web
0	3847	15298.0
1	3849	15296.0
2	3850	15300.0
3	4032	19814.0
4	4039	19815.0
5	4040	15303.0
6	4041	14975.0
7	4042	16042.0
8	4043	14980.0
9	4045	16041.0
10	4046	15269.0
11	4047	14977.0
12	4048	16044.0
13	4049	16043.0
14	4050	16449.0
15	4051	16045.0
16	4052	16030.0
17	4053	13127.0
18	4054	19816.0
19	4055	NaN

825 lignes

Des erreurs communes entre sku de web et id\_web de liaison.  
Même méthode de résolution que pour le fichier web

90 valeurs nulles dans id\_web

1 valeur mal écrite : « 14680-1 »

# Jointure ERP web

	product_id	id_web	onsale_web	price	stock_quantity	stock_status	sku	virtual	downloadable	rating_count	...
0	3847.0	15298.0	1.0	24.2	0.0	outofstock	15298.0	0.0	0.0	0.0	...
1	3849.0	15296.0	1.0	34.3	0.0	outofstock	15296.0	0.0	0.0	0.0	...
2	3850.0	15300.0	1.0	20.8	0.0	outofstock	15300.0	0.0	0.0	0.0	...
3	4032.0	19814.0	1.0	14.1	0.0	outofstock	19814.0	0.0	0.0	0.0	...
4	4039.0	19815.0	1.0	46.0	0.0	outofstock	19815.0	0.0	0.0	0.0	...

Jointure entre la tableau erp et web via tableau liaison (2 étapes)

Jointures externes pour garder le maximum de données

Pour vérifier la nouvelle table (jointureF) nous utilisons une colonne spéciale qui permet de vérifier l'adéquation des jointures

Nous ne gardons que les lignes qui on en commun une vente et un prix associé

714 lignes au final

# Analyse du CA

CA / produit

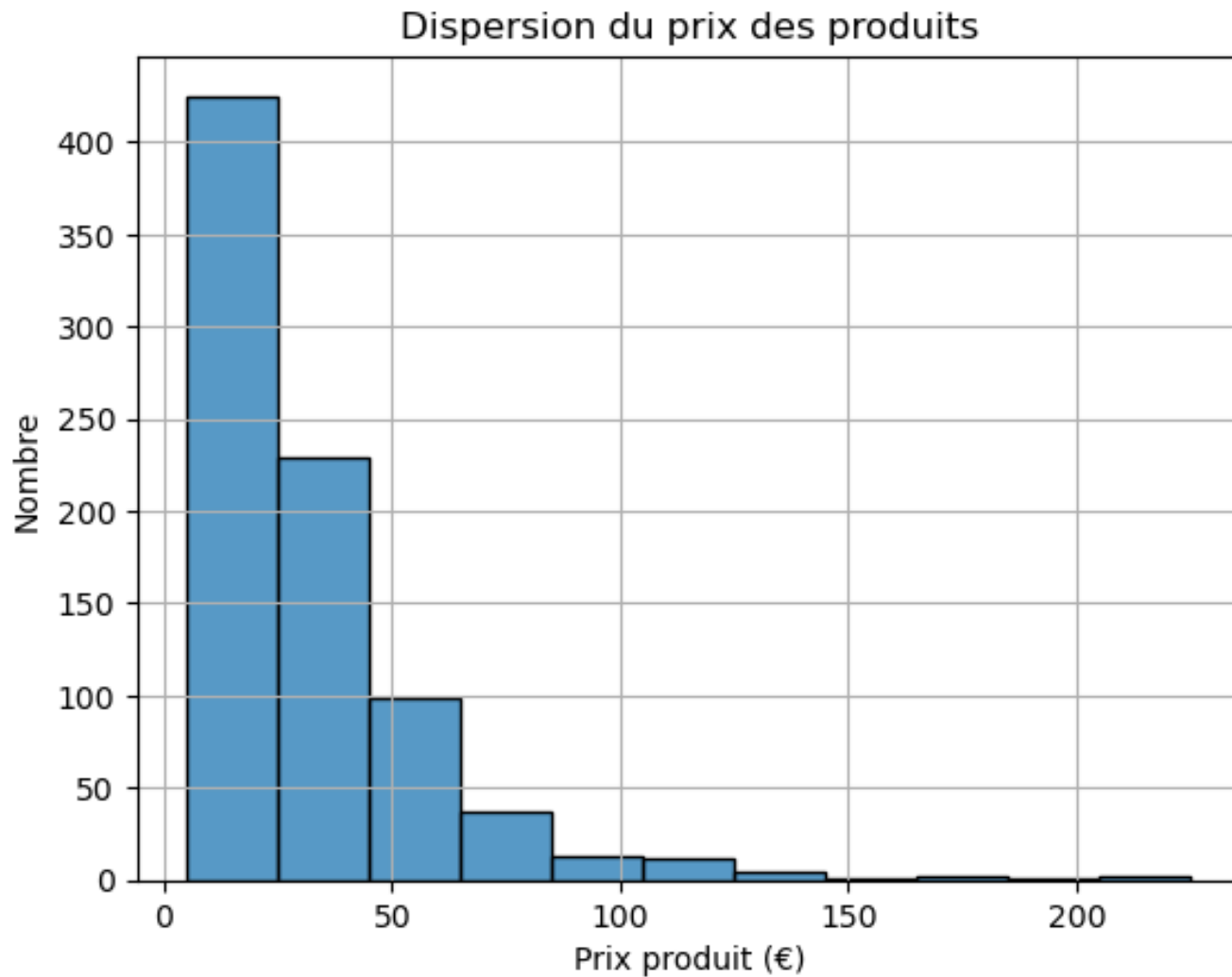
	product_id	price	total_sales	CA/produit
0	3847.0	24.2	6.0	145.2
1	3849.0	34.3	0.0	0.0
2	3850.0	20.8	0.0	0.0
3	4032.0	14.1	3.0	42.3
4	4039.0	46.0	0.0	0.0
5	4040.0	34.3	0.0	0.0
6	4041.0	32.7	0.0	0.0
7	4042.0	31.2	7.0	218.4
8	4043.0	60.0	3.0	180.0
9	4045.0	42.6	14.0	596.4

Top 5 des produits  
au plus fort CA

product_id	price	total_sales	CA/produit
4334.0	49.0	96.0	4704.0
4144.0	49.0	87.0	4263.0
4402.0	176.0	13.0	2288.0
4142.0	53.0	30.0	1590.0
4141.0	39.0	40.0	1560.0

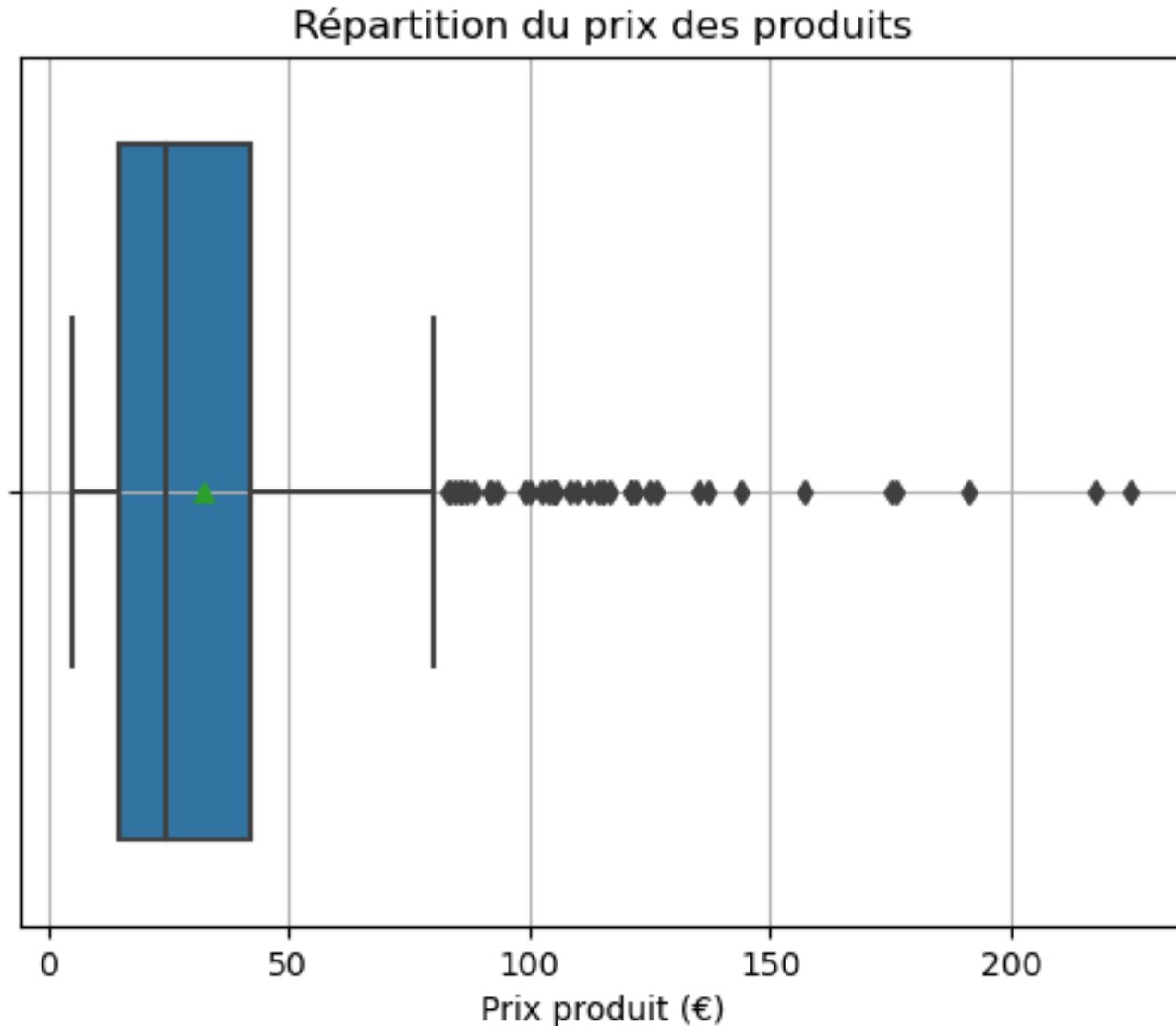
Le CA total en ligne est de  
70 568,6 euros

# Dispersion du prix des produits



Le prix des produits se concentre en dessous de 50 euros

# Répartition du prix des produits



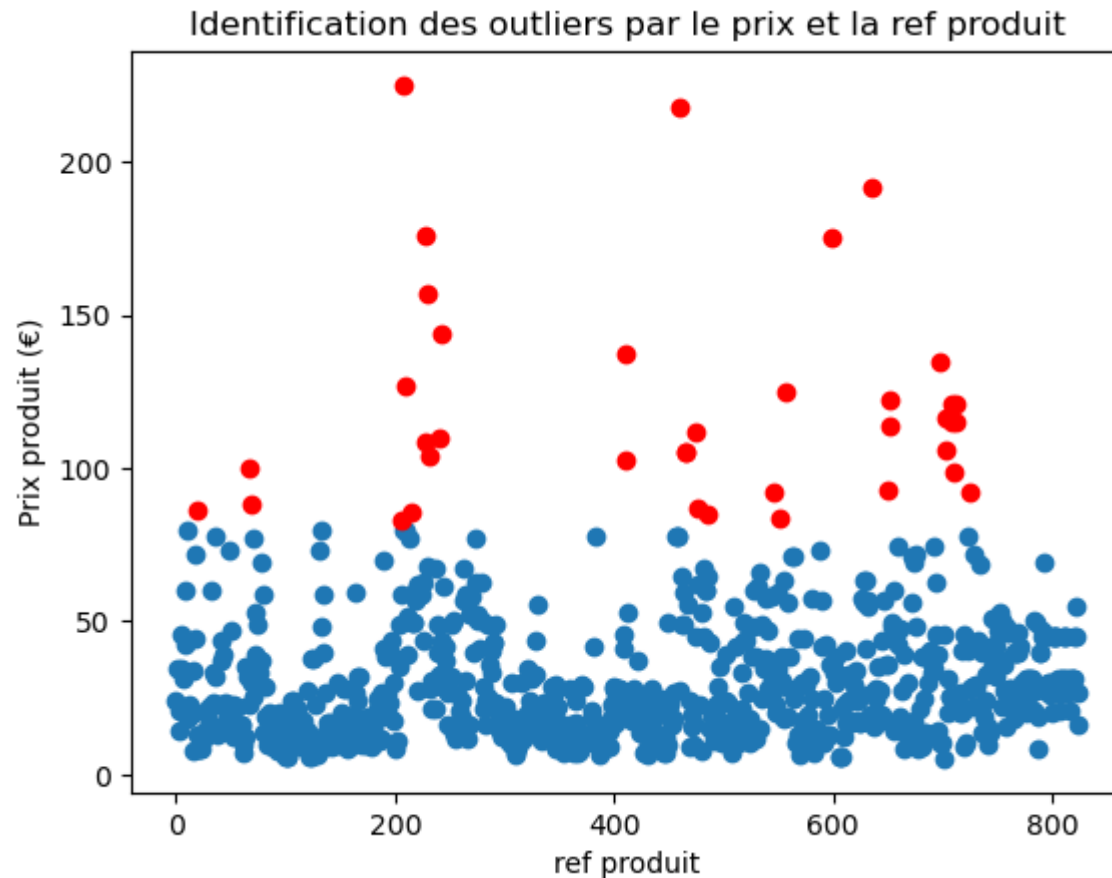
La médiane des prix est à  
**24,4 euros**

Le boxplot permet d'identifier visuellement les quartiles, la moyenne et de potentiels outliers

Les valeurs en dehors des 'moustaches' sont de potentiels outliers



# Répartition prix produits en fonction des ventes



En rouge, ce sont les outliers sur nos prix

On détermine ces outliers via plusieurs étapes :

- calcul du Q1 et Q3
- calcul de l'interquartile
- détermination des bornes sup et inf via

$$\text{borne\_inf} = q1 - 1.5 * \text{IQR}$$

$$\text{borne\_sup} = q3 + 1.5 * \text{IQR}$$

Les prix dépassant 83 euros ne peuvent être considérés comme des valeurs aberrantes, car les produits sont bien vendus