

Optimisation de la gestion des donnée de votre boutique

Les missions

- Nettoyage de la données
- Rapprochement des données de l'ERP et de WEB
- Calcule du chiffre d'affaire par produits
- Calcul du chiffre d'affaire total
- Analyse des prix

Les outils

Utilisation d'un Jupyter notebook qui est un logiciel permettant d'utiliser le langage de programmation Python.

Trois fichiers m'ont été transmis :

- Les données de l'ERP
- Les données du site WEB
- Une table de liaison

Nettoyage

WEB

Constat : lignes vides (NaN)

	sku	virtual	downloadable	rating_count	average_rating	total_sales	tax_status	tax_class	post_author
1	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
17	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
82	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
97	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
121	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
...
1378	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
1421	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
1448	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
1450	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN
1496	NaN	0	0	0	NaN	NaN	NaN	NaN	NaN

Suppression de ces lignes

Nettoyage

WEB

Constat : identifiants vides

	sku	virtual	downloadable	rating_count	average_rating	total_sales	tax_status	tax_class	post_author	post_date
353	NaN	0	0	0	0.0	3.0	taxable	NaN	2.0	2018-08-08 11:23:43
488	NaN	0	0	0	0.0	10.0	taxable	NaN	2.0	2018-07-31 12:07:23

J'ai trouvé deux identifiants vides avec tout de même des ligne produit non attribué.

Nettoyage

WEB

Constat : les doublons

sku	virtual	downloadable	rating_count	average_rating	total_sales	tax_status	tax_class	post_author	post_date
38	0	0	0	0.0	38.0	taxable	NaN	2.0	2018-04-18 12:25:58
38	0	0	0	0.0	38.0	NaN	NaN	2.0	2018-04-18 12:25:58
sku	tax_status				guid	post_type	post_mime_type		
15910	NaN	https://www.bottle-neck.fr/wp-content/uploads/...				attachment	image/jpeg		
15910	taxable	https://www.bottle-neck.fr/?post_type=product&...				product	NaN		

Après un filtrage sur les lignes en double, j'ai remarqué que l'une était la produit et l'autre le rattachement au site.

J'ai pris la décision de ne conservé que les lignes produits.

Nettoyage

ERP

Constat : identifiant en type texte

#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64

Modification

#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	object

Pour éviter toutes erreurs les identifiants doivent être en type texte.

Nettoyage

ERP

Constat : Vérifications des prix

Le prix max est 225.0 , le prix minimum est -8.0

	product_id	onsale_web	price	stock_quantity	stock_status
469	5017	0	-8.0	0	outofstock
739	6594	0	-1.0	192	instock

J'ai les prix maximum et minimum. Je vois un prix négatif et je filtre donc sur ces prix.

Il y a donc 2 prix négatifs mais avec aucune correspondance web.

Nettoyage

Liaisons

Constat : Nom de colonne

1	id_web	734 non-null	object
---	--------	--------------	--------

Modification

1	sku	734 non-null	object
---	-----	--------------	--------

Pour une unicité des nomenclature j'ai renommé la colonne "id_web" par "sku" comme dans le fichier web

Nettoyage

Liaisons

Constat : identifiant en type texte

#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64

Modification

#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	object

Pour éviter toutes erreurs les identifiants doivent être en type texte.

Nettoyage

Liaisons

Constat : “sku” vides

#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64
1	sku	734 non-null	object

On remarque qu’il manque des “sku” pour certain “product_id”. Je vais donc supprimer les “sku” vides (NaN).

Nettoyage

Liaisons

Constat : les "sku" avec des "-"

product_id	sku
4954	bon-cadeau-25-euros
7247	13127-1
7329	14680-1

Il y a 3 "sku" avec des tirets, après une rapide vérification ils semblent bien présent dans les données WEB. Il ne s'agit surment pas d'une erreur de frappe.

Jointure des tables ERP, liaisons et WEB

Jointure

	product_id	onsale_web	price	stock_quantity	stock_status	sku	virtual	downloadable	rating_count	average_rating	...	post_name	post_modified
0	3847	1	24.2	0	outofstock	15298	0.0	0.0	0.0	0.0	—	pierre-jean-villa-saint-joseph-preface-2018	2019-12-30 09:30:29
1	3849	1	34.3	0	outofstock	15296	0.0	0.0	0.0	0.0	—	pierre-jean-villa-saint-joseph-tilde-2017	2019-12-21 09:00:17
2	3850	1	20.8	0	outofstock	15300	0.0	0.0	0.0	0.0	—	pierre-jean-villa-croze-hermitage-accroche-coe...	2020-06-26 18:15:03
3	4032	1	14.1	0	outofstock	19814	0.0	0.0	0.0	0.0	—	pierre-jean-villa-igp-gamine-2018	2020-01-04 16:36:01
4	4039	1	46.0	0	outofstock	19815	0.0	0.0	0.0	0.0	—	pierre-jean-villa-cote-rotie-carmina-2017	2020-01-04 16:36:10
...	—	—	—	—	—	—	—	—	—	—	—	—	—
911	7023	1	27.5	15	instock	15891	0.0	0.0	0.0	0.0	—	camin-larredya-jurancon-sec-la-virada-2018	2020-08-26 17:35:02
912	7025	1	69.0	2	instock	15887	0.0	0.0	0.0	0.0	—	jamet-cote-rotie-fructus-voluptas-2018	2020-08-14 18:15:03
913	7247	1	54.8	23	instock	13127-1	0.0	0.0	0.0	0.0	—	clos-du-mont-olivet-chateauneuf-du-pape-2007-2	2020-07-20 17:09:06
914	7329	0	26.5	14	instock	14680-1	NaN	NaN	NaN	NaN	—	NaN	NaN
915	7338	1	16.3	45	instock	16230	0.0	0.0	0.0	0.0	—	domaine-saint-nicolas-fiefs-vendeens-blanc-les...	2020-08-13 10:45:03

734 rows × 33 columns

Le chiffres d'affaires produits

Les Analyses

J'ai créée la colonne "CA" pour avoir le chiffre d'affaire par produits.

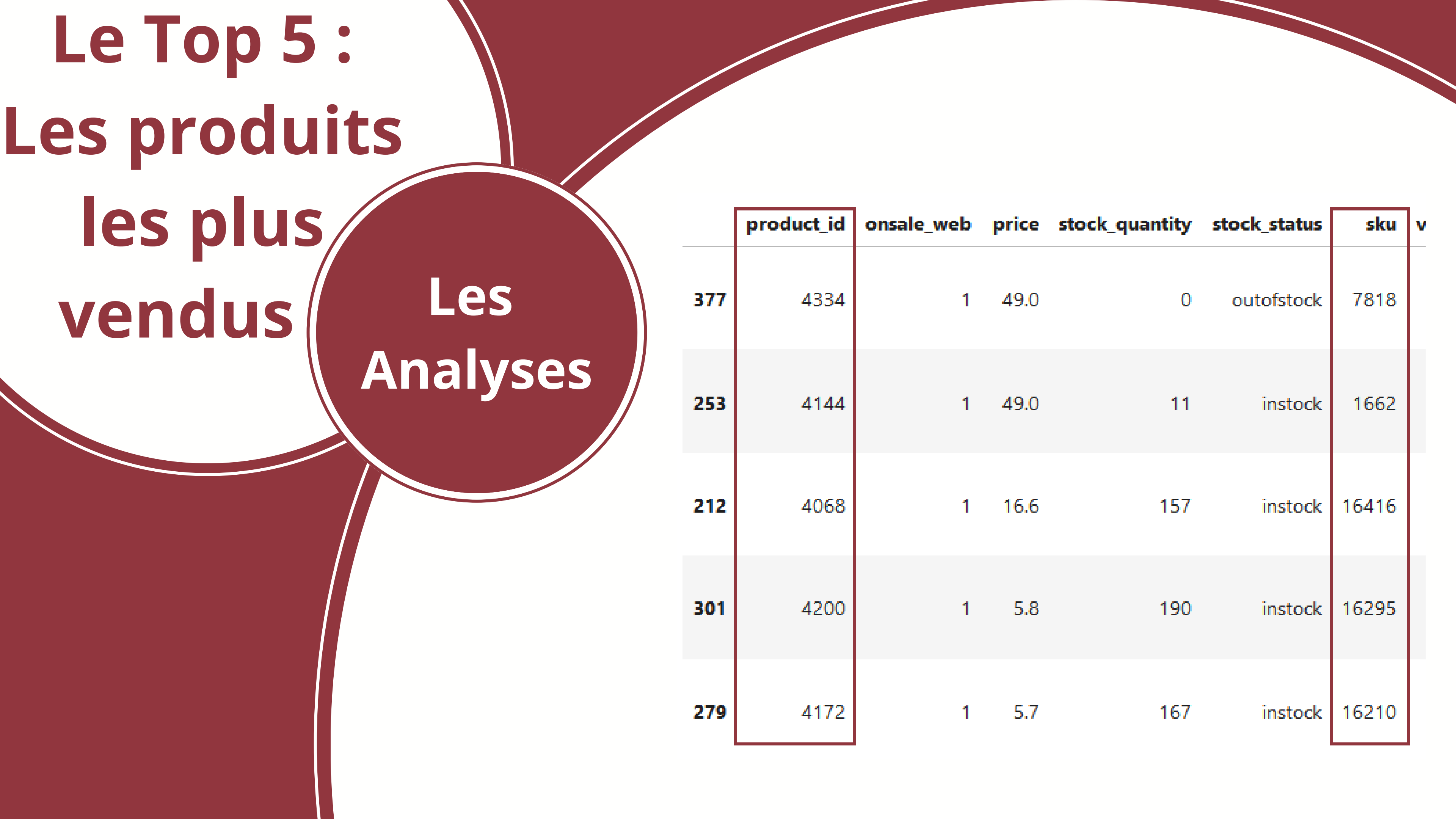
CA
145.2
0.0

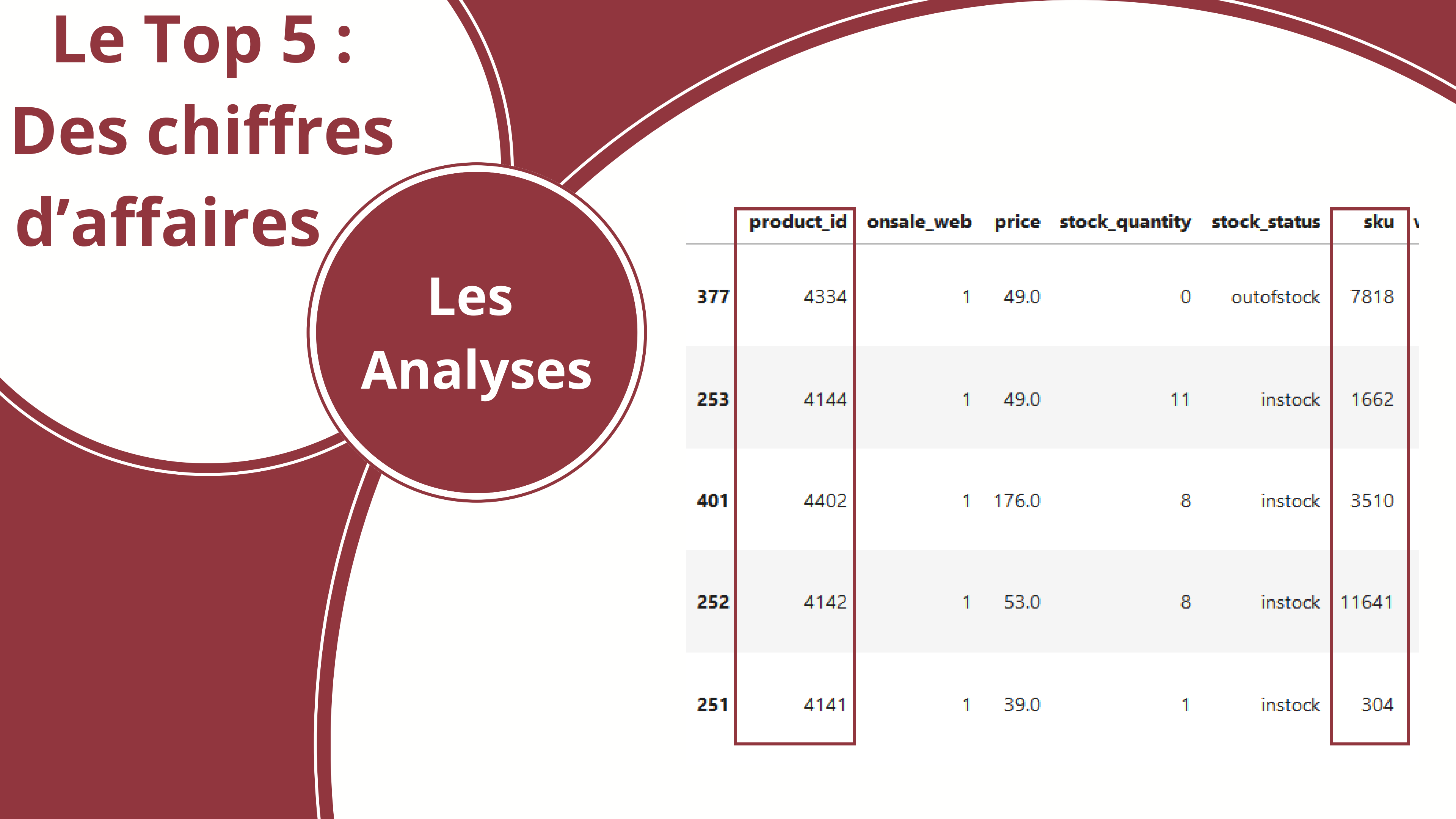
"CA" = "prix" * "total des ventes"

**Le chiffres
d'affaires
global**

**Les
Analyses**

70 568.6 €





Le Top 5 : Des chiffres d'affaires

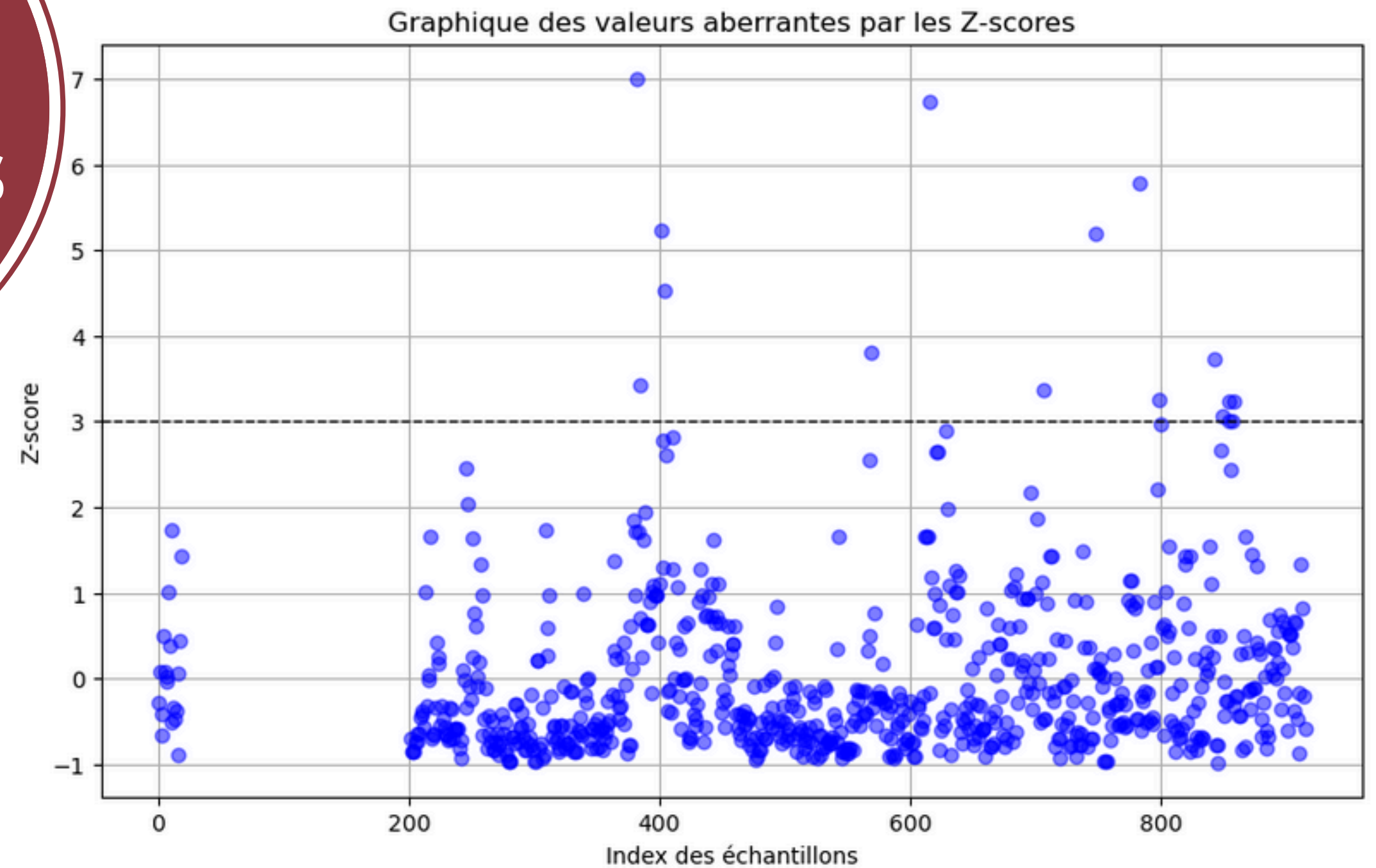
Les
Analyses

	product_id	onsale_web	price	stock_quantity	stock_status	sku
377	4334	1	49.0	0	outofstock	7818
253	4144	1	49.0	11	instock	1662
401	4402	1	176.0	8	instock	3510
252	4142	1	53.0	8	instock	11641
251	4141	1	39.0	1	instock	304

Méthode des Z - score

Les
outliers

Le graphique



Méthode des Z - score

Les
outliers

La liste des 16 valeurs aberrantes trouvées

	product_id	onsale_web	price	stock_quantity	stock_status	sku
382	4352	1	225.0	0	outofstock	15940
616	5001	1	217.5	20	instock	14581
783	5892	1	191.3	10	instock	14983
401	4402	1	176.0	8	instock	3510
748	5767	1	175.0	12	instock	15185
404	4406	1	157.0	3	instock	7819
569	4904	1	137.0	13	instock	14220
843	6126	1	135.0	10	instock	14923
384	4355	1	126.5	2	instock	12589
706	5612	1	124.8	12	instock	14915

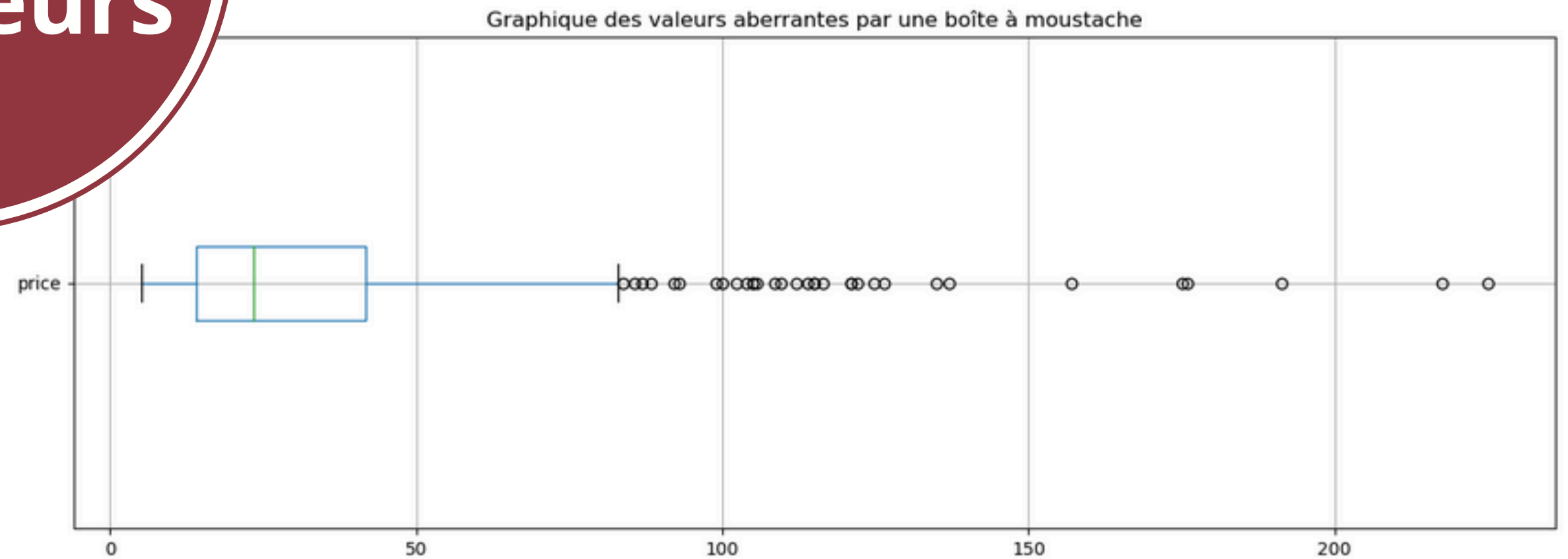
	product_id	onsale_web	price	stock_quantity	stock_status	sku
799	5917	1	122.0	4	instock	14775
855	6213	1	121.0	7	instock	15072
858	6216	1	121.0	6	instock	15070
849	6202	1	116.4	14	instock	15126
854	6212	1	115.0	2	instock	13996
857	6215	1	115.0	4	instock	12790

16 rows × 34 columns

Méthode des boxplot

Les outliers

Le graphique



Méthode des boxplot

Les
outliers

La liste des 32 valeurs aberrantes trouvées

	product_id	onsale_web	price	stock_quantity	stock_status	sku
382	4352	1	225.0	0	outofstock	15940
616	5001	1	217.5	20	instock	14581
783	5892	1	191.3	10	instock	14983
401	4402	1	176.0	8	instock	3510
748	5767	1	175.0	12	instock	15185
404	4406	1	157.0	3	instock	7819
569	4904	1	137.0	13	instock	14220
843	6126	1	135.0	10	instock	14923
384	4355	1	126.5	2	instock	12589
706	5612	1	124.8	12	instock	14915
799	5917	1	122.0	4	instock	14775
855	6213	1	121.0	7	instock	15072
858	6216	1	121.0	6	instock	15070
849	6202	1	116.4	14	instock	15126
854	6212	1	115.0	2	instock	13996
857	6215	1	115.0	4	instock	12790

	product_id	onsale_web	price	stock_quantity	stock_status	sku
800	5918	1	114.0	8	instock	14773
629	5025	1	112.0	0	outofstock	13914
411	4582	1	109.6	7	instock	12857
402	4404	1	108.5	2	instock	3507
848	6201	1	105.6	7	instock	14596
621	5007	1	105.0	17	instock	12791
622	5008	1	105.0	10	instock	11602
405	4407	1	104.0	6	instock	3509
568	4903	1	102.3	20	instock	14805
245	4115	1	100.0	11	instock	15382
856	6214	1	99.0	7	instock	11601
798	5916	1	93.0	3	instock	14774
696	5565	1	92.0	0	outofstock	19822
247	4132	1	88.4	5	instock	11668
630	5026	1	86.8	2	instock	13913
388	4359	1	85.6	0	outofstock	13853

32 rows × 34 columns

Conclusions

Pour conclure, j'ai effectué des recherches sur les prix des bouteilles dont les prix sont ressortis, et ils correspondent aux prix du marché.

Ces valeurs ne nécessitent pas d'être exclues car il ne semble pas y avoir d'erreur dans la saisie de leurs prix.

Merci
pour votre
attention