

# Optimisez la gestion & nettoyez les données du stock d'une boutique

AH-SEE Benoit

Business Intelligence Analyst

10 mai 2024

# Analyses Exploratoires des Données

---

- 3 datasets
  - ❖ 1) Fichier Erp : 825 observations X 6 colonnes
  - ❖ 2) Fichier Web : 1 513 observations x 29 colonnes
  - ❖ 3) Fichier Liaison : 825 observations X 2 colonnes
- Traitement réalisé
  - ❖ Nettoyage des données avant l'analyse avec Python sur Jupyter Notebook



# Analyses Exploratoires des Données


- *Nettoyage du 1<sup>er</sup> fichier « ERP » => Stock*

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	stock_status_2
4	4039	1	46.0	3	outofstock	23.77	instock
398	4885	1	18.7	0	instock	9.66	outofstock
449	4973	0	10.0	-10	outofstock	4.96	instock
573	5700	1	44.5	-1	outofstock	22.30	instock



AVANT

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	stock_status_2
4	4039	1	46.0	3	instock	23.77	instock
398	4885	1	18.7	0	outofstock	9.66	outofstock
449	4973	0	10.0	0	outofstock	4.96	outofstock
573	5700	1	44.5	0	outofstock	22.30	outofstock




APRES

Transformation des données incohérentes

# Analyses Exploratoires des Données


- *Nettoyage du 1<sup>er</sup> fichier « ERP » => Prix de vente*

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	stock_status_2
151	4233	0	-20.0	0	outofstock	10.33	outofstock
469	5017	0	-8.0	0	outofstock	4.34	outofstock
739	6594	0	-9.1	19	instock	4.61	instock



AVANT

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	stock_status_2
151	4233	0	20.0	0	outofstock	10.33	outofstock
469	5017	0	8.0	0	outofstock	4.34	outofstock
739	6594	0	9.1	19	instock	4.61	instock



APRES

Transformation des données incohérentes

# Analyses Exploratoires des Données

- *Nettoyage du 2<sup>ème</sup> fichier « Web »*

Lignes avec des valeurs non-entières dans la colonne 'sku':

	sku	total_sales	tax_status	post_author	post_date	post_date_gmt	product_type	post_title
8	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN
20	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN
30	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN
37	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN
41	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN
...	...	...	...	...	...	...	...	...
1387	bon-cadeau-25-euros	7.0	taxable	1.0	2018-06-01 13:53:46	2018-06-01 11:53:46	NaN	Bon cadeau de 25€
1429	NaN	NaN	NaN	NaN	NaT	NaT	NaN	NaN

## Suppression

- 1) 11 colonnes vides
- 2) 2 codes article incohérents : « bon cadeau-25-euros » et « 13127-1 »
- 3) 85 lignes « NaN » de la colonne sku vides
- 4) 712 lignes avec code article en doublon dans la colonne sku

1 513 lignes x 29 colonnes => 712 lignes X 18 colonnes, soit environ 50 % de « nettoyage »

# Analyses Exploratoires des Données

- *Nettoyage du 3<sup>ème</sup> fichier « Liaison »*

	id_web	product_id
19	NaN	4055
49	NaN	4090
50	NaN	4092
119	NaN	4195
131	NaN	4209
...	...	...
817	NaN	7196
818	NaN	7200
819	NaN	7201
820	NaN	7203
821	NaN	7204

Lignes avec des valeurs non-entières

	id_web	product_id
443	bon-cadeau-25-euros	4954
822	13127-1	7247
823	14680-1	7329

## Suppression :

- 1) 91 lignes sans correspondance
- 2) 3 lignes avec code art. incohérent :  
« bon-cadeau-25-euros », « 13127-1 »  
et « 14680-1 »



825 lignes x 2 colonnes => 731 lignes X 2 colonnes, soit environ 12 % de « nettoyage »

# Fusion ou consolidation des données

- *Fusion des fichiers « Erp » et « Liaison »*

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	id_web
0	3847	1	24.2	16	instock	12.88	15298
1	3849	1	34.3	10	instock	17.54	15296
2	3850	1	20.8	0	outofstock	10.64	15300
3	4032	1	14.1	26	instock	6.92	19814
4	4039	1	46.0	3	instock	23.77	19815

1

- *Fusion des fichiers « Erp et Liaison » et « Web »*

	product_id	onsale_web	price	stock_quantity	stock_status	purchase_price	id_web	sku	total_sales	tax_status	post_author
0	3847	1	24.2	16	instock	12.88	15298	15298	6.0	taxable	2.0
1	3849	1	34.3	10	instock	17.54	15296	15296	9.0	taxable	2.0

2

Fichier final : 712 lignes X 25 colonnes et conforme au RGPD

# Analyses univariées du prix

- *Moyenne : 32,21 € ( 23 006 / 712 )*

Mesure statistique représentant la valeur centrale d'un ensemble de données

- *Ecart-type : 27,62 €*

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Mesure de dispersion statistique indiquant à quel point les prix sont éloignées de la moyenne.  
Dans notre cas, écart-type inférieur à la moyenne, donc dispersion plus faible.

- *Seuil de prix avec Z-score supérieur à 3 : 115,17 € ( ( 115,17-32,21)/27,62 )*

$$Z = \frac{(X - \mu)}{\sigma}$$

Également appelé score standardisé, c'est une mesure statistique qui indique à quelle distance une valeur donnée se situe par rapport à la moyenne d'un ensemble de données, en termes d'écart-type.

- ❖ Z-score positif : valeur au-dessus de la moyenne.
- ❖ Z-score négatif : valeur en-dessous de la moyenne.



# Analyses univariées du prix

## ➤ *Outlier par l'interval interquartile*

- ❖ L'outlier, ou valeur aberrante, est une observation qui se situe à une distance significativement différente des autres observations dans un ensemble de données.
- ❖ Elles peuvent avoir un impact disproportionné sur les résultats statistiques.
- ❖ Il est important de les identifier et de décider s'ils doivent être traités ou supprimés lors de l'analyse des données.

```
#Définissez un seuil pour les articles "outliers" en prix
# Calculer le premier et le troisième quartile
Q1 = df_merge['price'].quantile(0.25)
Q3 = df_merge['price'].quantile(0.75)

# Calculer l'intervalle interquartile (IQR)
IQR = Q3 - Q1

# Définir les bornes pour les outliers
seuil_inf = Q1 - 3 * IQR
seuil_sup = Q3 + 3 * IQR
```

Nombre d'outliers en prix : 8

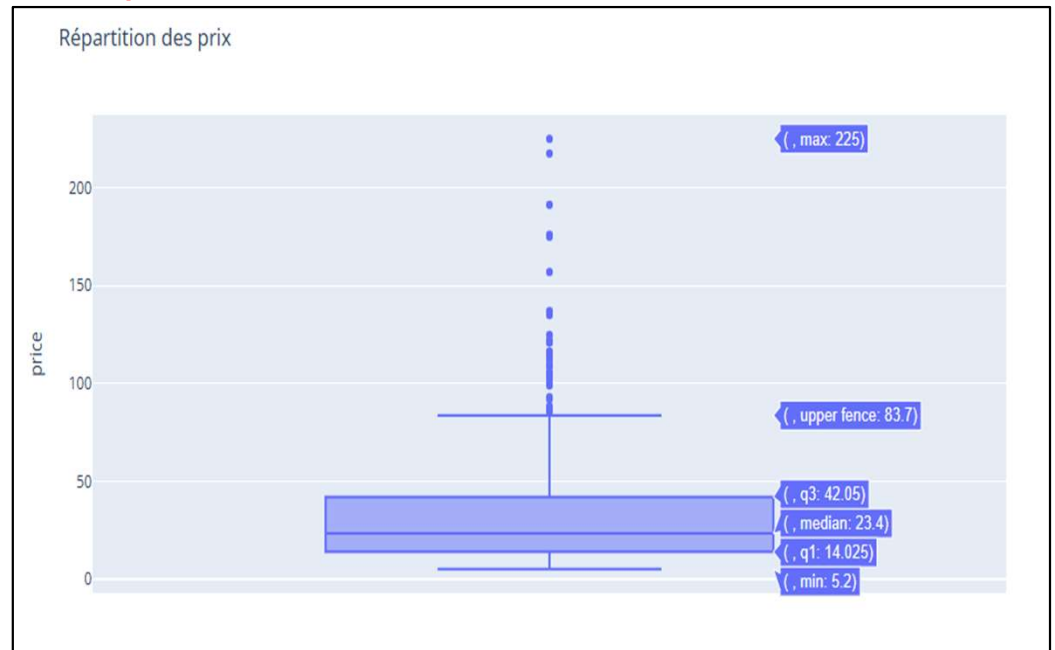
	product_id	post_title	price
641	6126	Champagne Gosset Célébris Vintage 2007	135.0
381	4904	Domaine Des Croix Corton Charlemagne Grand Cru...	137.0
221	4406	Cognac Frapin Château de Fontpinot 1989 20 Ans...	157.0
552	5767	Camille Giroud Clos de Vougeot 2016	175.0
218	4402	Cognac Frapin VIP XO	176.0
586	5892	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	191.3
425	5001	David Duband Charmes-Chambertin Grand Cru 2014	217.5
199	4352	Champagne Egly-Ouriet Grand Cru Millésimé 2008	225.0

# Analyses univariées du prix

- *Visualisation de la dispersion des prix : « Boîte à moustache »*



Pandas



Plotly express

# Analyses univariées du prix

- *Outlier : Champagne Egly-Ouriet – Grand Cru - Millésime 2008*



**Champagne EGLY-OURIET - Grand Cru - Millésime 2008**  
Champagne

Champagne EGLY-OURIET - Grand Cru - Millésime 2008  
La maison Egly-Ouriet est née en 1945 et est passée de quelques hectares de vignes à plus de 9 hectares aujourd'hui. Si l'on devait résumer la philosophie de la maison, citait Francis Egly : « Les vignes sont comme les chevaux : si on veut les contrôler, il faut les faire vivre. » Un producteur et cavalier plein de bon sens qui cultive avec passion des vignes issues de sélections de masse exceptionnelles. Les vins sont vinifiés en barriques (depuis 1995) et sans fermentation malolactique pour leur donner plus de fraîcheur.

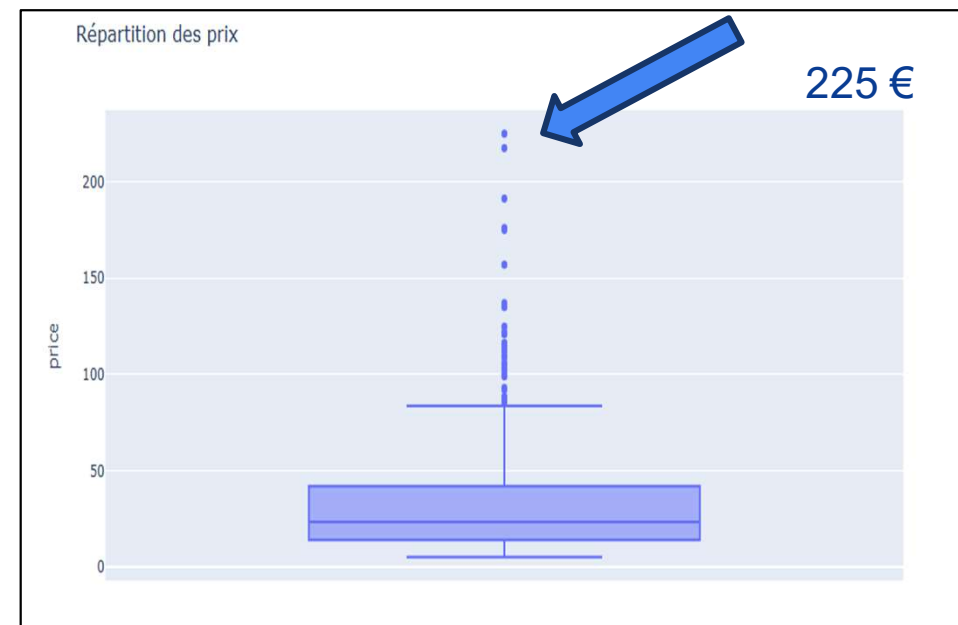
Le Grand Cru Millésime 2008 est un assemblage de 70% de Pinot Noir et de 30% de Chardonnay. Pour certains des meilleurs vignerons de Champagne. Ce millésime exceptionnel 2008, ce millésime sera très peu donné pour laisser s'exprimer la meilleure qualité du vin. C'est un champagne à robe dorée et aux bulles très fines et douces. Un vin avant d'être un champagne. La bouche est typique des vins du domaine, avec beaucoup de profondeur et une grande minéralité. Un millésime à déguster une fois dans sa vie !

Unité **199.00 €** Taxes incluses  
165.83 € HT

**RUPTURE DE STOCK**

**ALERTE DE STOCK** Réserve aux membres

Google



Boîte à moustache

## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

Les 20 premiers articles en termes de chiffre d'affaires sont :

	product_id	post_title	ca_par_article
0	4150	Champagne Mailly Grand Cru Intemporelle 2010	6844.0
1	4352	Champagne Egly-Ouriet Grand Cru Millésimé 2008	2475.0
2	4726	François Baur Pinot Noir Schlittweg 2017	1549.4
3	5067	Albert Mann Pinot Noir Grand H 2017	1317.8
4	5379	Argentine Mendoza Alamos Torrontes 2017	1232.1
5	5892	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	1147.8
6	4353	Champagne Egly-Ouriet Grand Cru Brut Rosé	1113.0
7	5826	Agnès Levet Côte Rôtie Améthyste 2017	824.0
8	6212	Domaine des Comtes Lafon Volnay 1er Cru Santen...	805.0
9	5026	Champagne Agrapart & amp; Fils Minéral Extra Br...	781.2
10	5008	Domaine des Comtes Lafon Volnay 1er Cru Santen...	735.0
11	5767	Camille Giroud Clos de Vougeot 2016	700.0
12	6126	Champagne Gosset Célébris Vintage 2007	675.0
13	5025	Champagne Agrapart & amp; Fils L'Avizoise Extra...	672.0
14	6201	David Duband Chambolle-Musigny 1er Cru Les Sen...	633.6
15	4406	Cognac Frapin Château de Fontpinot 1989 20 Ans...	628.0
16	4647	Bernard Baudry Chinon Rouge La Croix Boissée 2017	627.0
17	4358	Champagne Larmandier-Bernier Grand Cru Vieille...	616.0
18	4359	Champagne Larmandier-Bernier Grand Cru Les Che...	599.2
19	6214	Domaine des Comtes Lafon Volnay 1er Cru Champa...	594.0

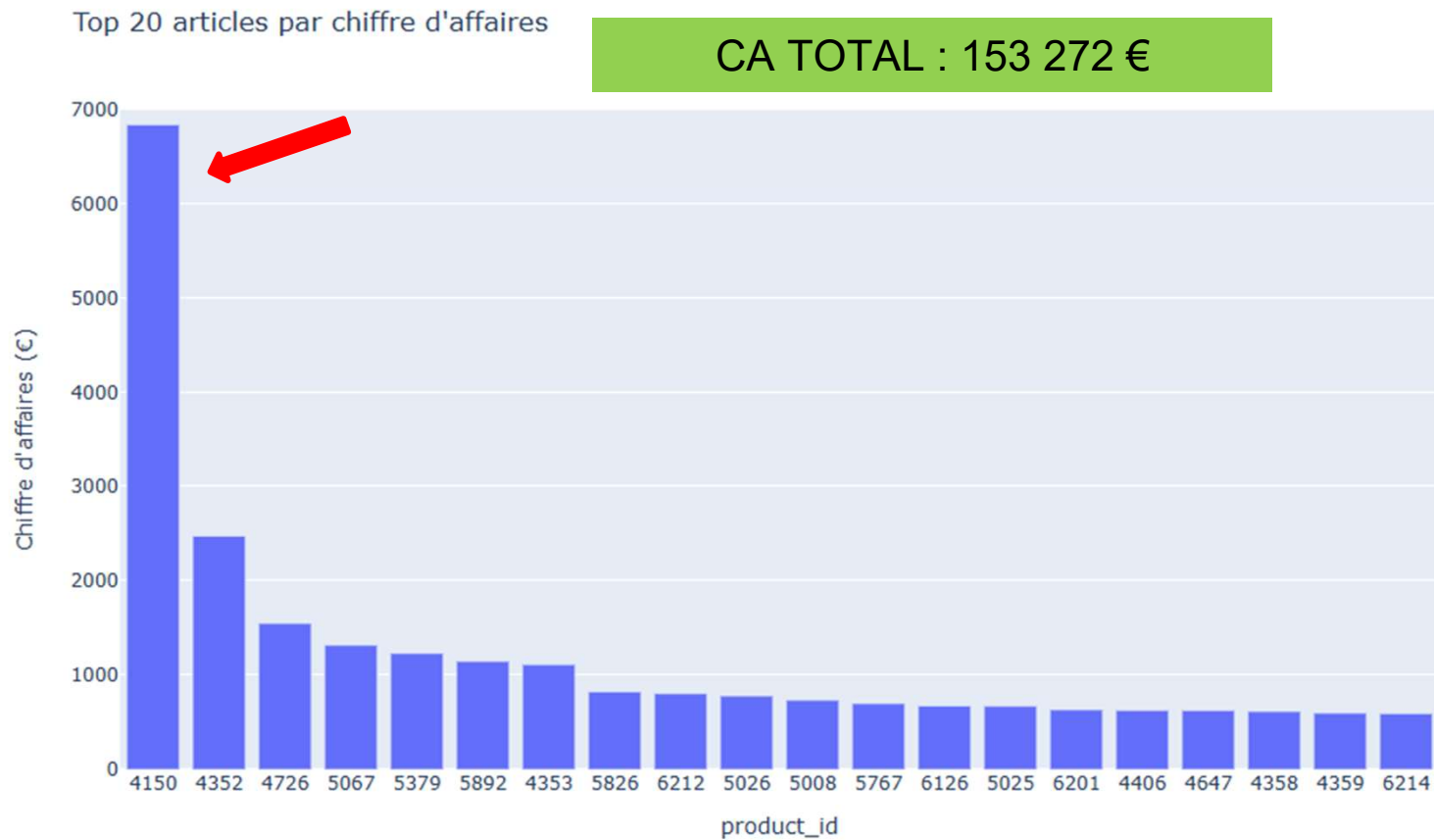


	product_id	post_title	ca_par_article	ca_part	ca_cumulative
0	4150	Champagne Mailly Grand Cru Intemporelle 2010	6844.0	0.044652	0.044652
1	4352	Champagne Egly-Ouriet Grand Cru Millésimé 2008	2475.0	0.016148	0.060800
2	4726	François Baur Pinot Noir Schlittweg 2017	1549.4	0.010109	0.070909
3	5067	Albert Mann Pinot Noir Grand H 2017	1317.8	0.008598	0.079507
4	5379	Argentine Mendoza Alamos Torrontes 2017	1232.1	0.008039	0.087545
...	...	...	...	...	...
414	4085	Rimauresq Côtes de Provence Rouge Cru Classé 2017	152.0	0.000992	0.796015
415	4889	Saumaize-Michelin Pouilly-Fuissé Vignes Blanch...	151.8	0.000990	0.797006
416	4194	Château de La Liquière Faugères 2017	151.2	0.000986	0.797992
417	5962	Camin Larredya Jurançon Moelleux Au Capcéu 2018	150.0	0.000979	0.798971
418	4183	Borie La Vitarèle Saint-Chinian Les Crès 2016	149.8	0.000977	0.799948
419 rows x 5 columns					



Affichage des 20 meilleures ventes en CA  
59 % des articles représentent 80 % du CA.

## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



CA : 6 844 €  
Qté : 116 UV





## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

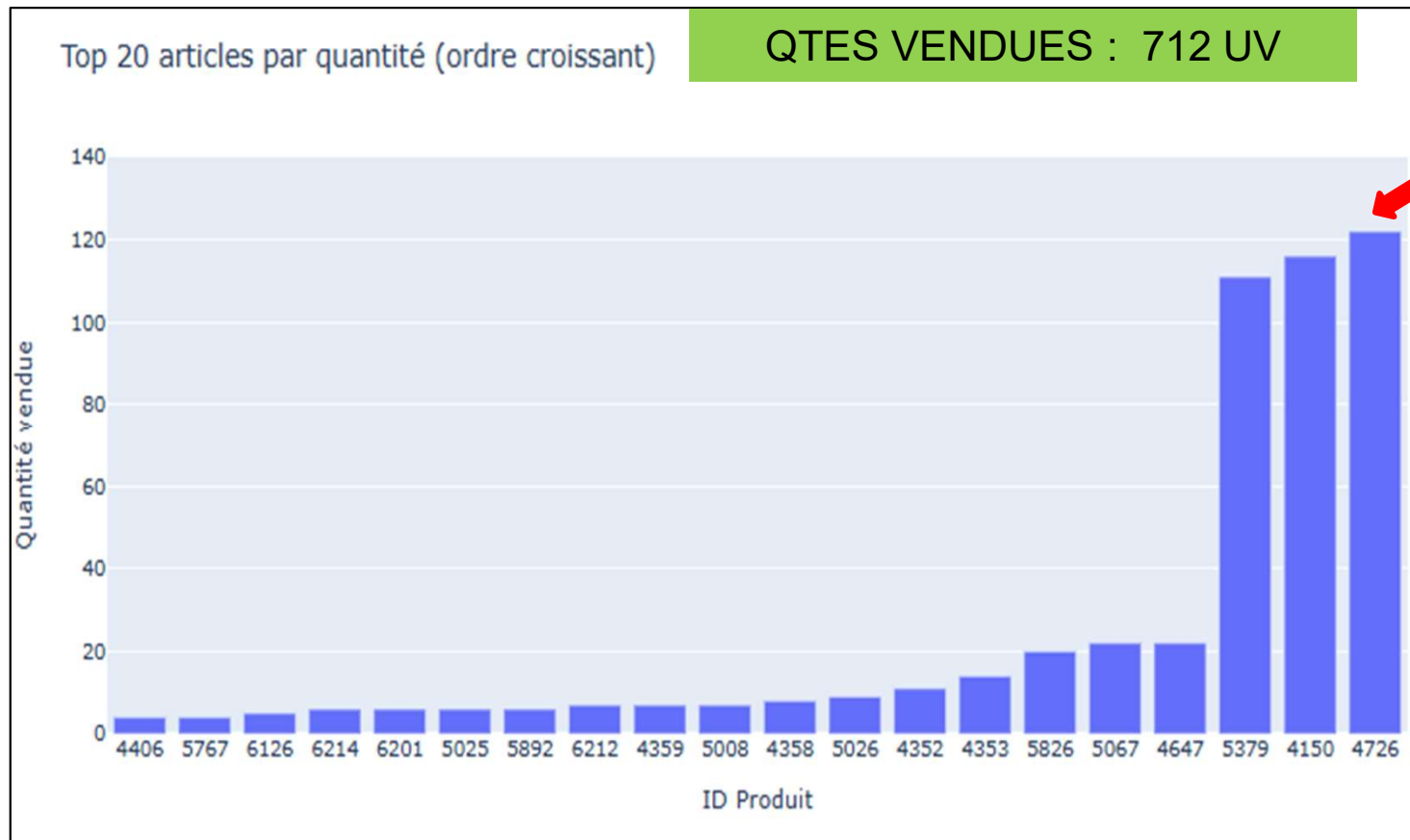
Les 20 premiers articles en termes de quantité vendue sont :

	product_id	post_title	total_sales
0	4726	François Baur Pinot Noir Schlittweg 2017	122.0
1	4150	Champagne Mailly Grand Cru Intemporelle 2010	116.0
2	5379	Argentine Mendoza Alamos Torrontes 2017	111.0
3	4867	Château De La Selve IGP Coteaux de l'Ardèche M...	36.0
4	4203	Mas Laval IGP Pays d'Hérault Les Pampres Blanc...	27.0
5	4275	I Fabbri Chianti Classico Lamole 2017	24.0
6	5067	Albert Mann Pinot Noir Grand H 2017	22.0
7	4647	Bernard Baudry Chinon Rouge La Croix Boissée 2017	22.0
8	6129	Moulin de Gassac IGP Pays d'Hérault Guilhem Bl...	20.0
9	5826	Agnès Levet Côte Rôtie Améthyste 2017	20.0
10	4220	Xavier Frissant Touraine Amboise Chenin Les Pi...	18.0
11	6569	Decelle-Villa Chorey-Lès-Beaune 2016	17.0
12	5803	Château Tour Haut-Caussan Médoc 2015	17.0
13	5778	Maurel Pays d'Oc Merlot 2018	17.0
14	4870	Triennes IGP Méditerranée Rosé 2019	16.0
15	4105	Emile Boeckel Sylvaner Vieilles Vignes 2019	16.0
16	4059	Mourgues du Grès Costières de Nîmes Galets Ros...	16.0
17	4863	Château Ollieux Romanis Corbières Rosé Classiq...	16.0
18	5777	Maurel Pays d'Oc Cabernet-Sauvignon 2017	16.0
19	5695	Philippe Nusswitz IGP Cévènnès Rosé O Pale 2019	16.0

	product_id	post_title	total_sales	part_quantite	cumulative_part_quantite
321	4726	François Baur Pinot Noir Schlittweg 2017	122.0	0.020099	0.020099
76	4150	Champagne Mailly Grand Cru Intemporelle 2010	116.0	0.019110	0.039209
449	5379	Argentine Mendoza Alamos Torrontes 2017	111.0	0.018287	0.057496
365	4867	Château De La Selve IGP Coteaux de l'Ardèche M...	36.0	0.005931	0.063427
122	4203	Mas Laval IGP Pays d'Hérault Les Pampres Blanc...	27.0	0.004448	0.067875
...	...	...	...	...	...
652	6212	Domaine des Comtes Lafon Volnay 1er Cru Santen...	7.0	0.001153	0.796211
657	6221	Planeta Sicilia Eruzione 1614 2017	7.0	0.001153	0.797364
245	4615	Alphonse Mellot Sancerre Blanc La Moussière 2018	7.0	0.001153	0.798517
244	4614	Gilbert Picq Chablis En Vaudécorse 2017	7.0	0.001153	0.799671
183	4287	Domaine Hauvette Les Baux de Provence Cornalin...	7.0	0.001153	0.800824
424 rows x 5 columns					

Affichage des 20 meilleures ventes en terme de quantité  
59 % des articles représentent 80 % du CA.

## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



Unités vendues : 122  
CA : 1 549 € 40



## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

'Top 20 des produits avec le plus de mois de stock :'

	product_id	post_title	stock_rotation
70	4142	Champagne Gosset Grand Millésime 2006	31.25
641	6126	Champagne Gosset Célébris Vintage 2007	27.60
202	4356	Champagne Egly-Ouriet Premier Cru Les Vignes d...	27.00
197	4348	Champagne Egly-Ouriet Grand Cru Brut Tradition	25.00
74	4148	Champagne Mailly Grand Cru Brut Rosé	23.67
203	4357	Champagne Larmandier-Bernier Latitude	23.00
71	4144	Champagne Gosset Grand Rosé	22.75
436	5025	Champagne Agrapart & Fils L'Avizoise Extra...	22.67
198	4350	Champagne Egly-Ouriet Grand Cru Extra Brut V.P.	20.71
194	4334	Champagne Gosset Grand Blanc de Blancs	20.29
75	4149	Champagne Mailly Grand Cru Intemporelle Rosé 2009	20.20
510	5612	Domaine Weinbach Gewurztraminer Grand Cru Furs...	19.00
227	4582	Château de Meursault Puligny-Montrachet 1er Cr...	18.00
435	5024	Champagne Agrapart & Fils Les 7 Crus Brut ...	17.17
414	4970	Champagne Larmandier-Bernier Terre de Vertus P...	16.67
586	5892	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	16.33
205	4359	Champagne Larmandier-Bernier Grand Cru Les Che...	16.00
69	4141	Champagne Gosset Grande Réserve	15.38
72	4146	Champagne Mailly Grand Cru Brut Réserve	14.33
438	5027	Champagne Agrapart & Fils Terroirs Brut Bl...	13.17

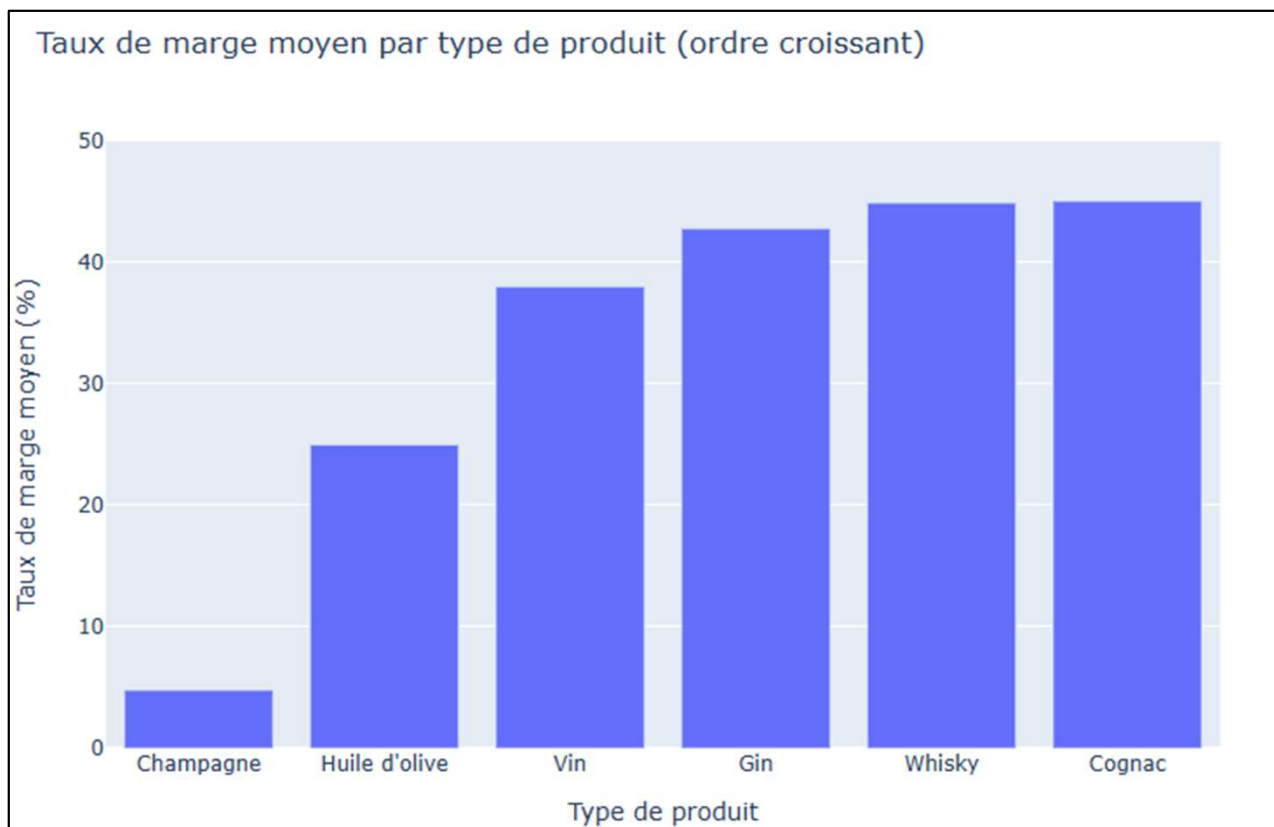
Flop 20 articles par mois de stock



Stock : 125 unités  
Ventes : 4 unités  
Rotation : 31 mois



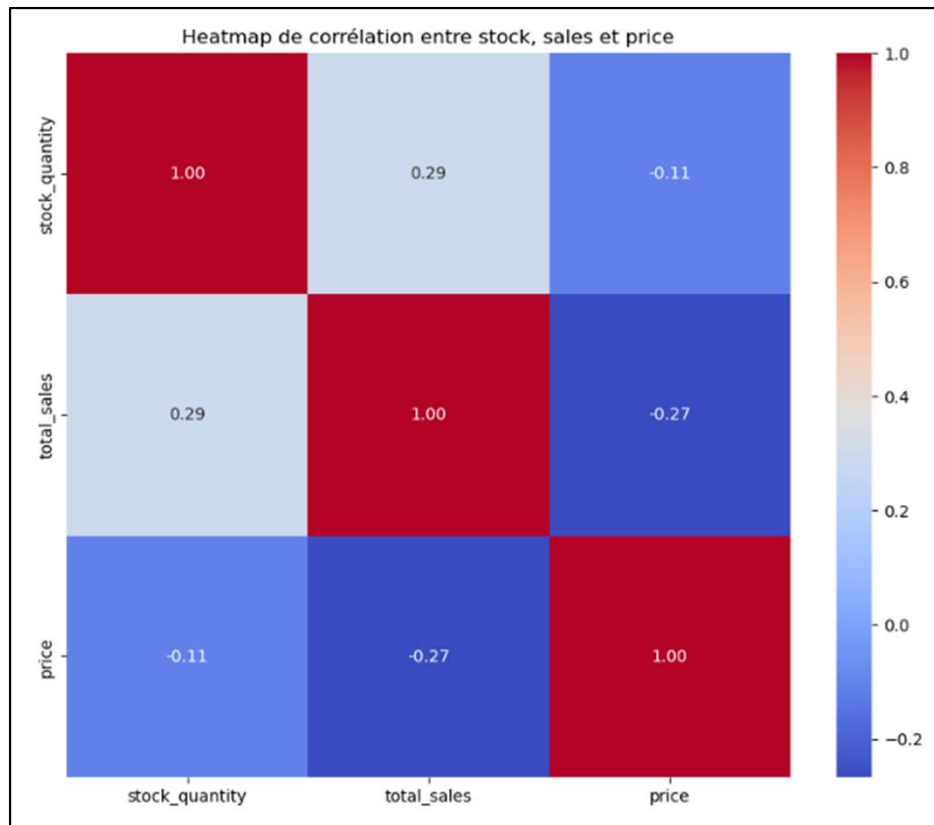
## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



	product_type	taux_marge
0	Champagne	4.79
1	Cognac	45.07
2	Gin	42.80
3	Huile d'olive	25.00
4	Vin	38.01
5	Whisky	44.92



## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations



### Analyse des corrélations :

1) Stock\_quantity / Total\_sales : 0,29

➤ Corrélation positive modérée

2) Price / Total Sales : - 0,27

➤ Corrélation négative modérée

3) Stock\_quantity / Price : - 0,11

➤ Pas de corrélation claire

### Limites de l'analyse :

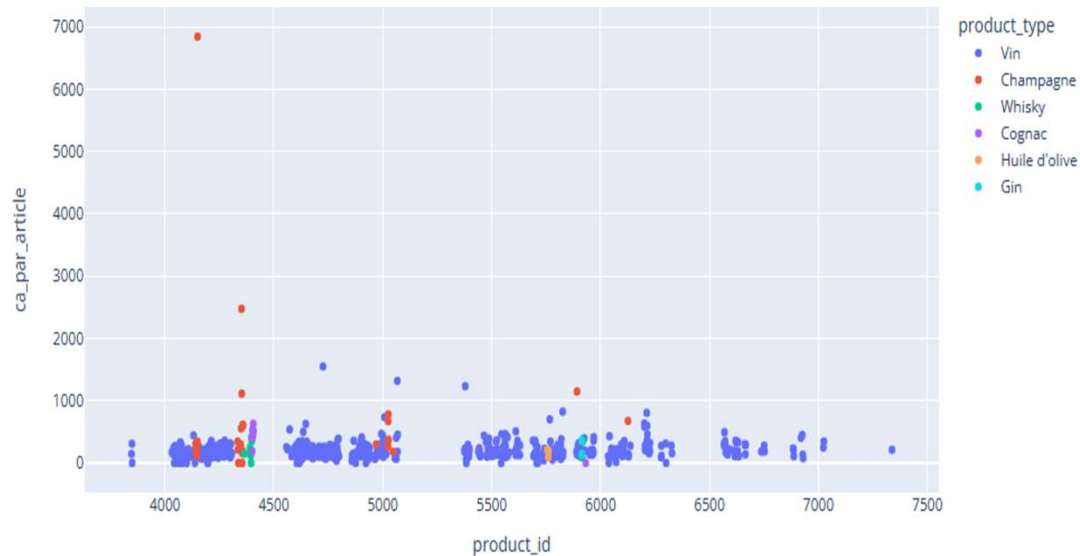
- ❖ Un seul mois de données : octobre
- ❖ Des données incohérentes
- ❖ Pas de comparatif N/N-1
- ❖ Manque le nombre de clients

# Analyses complémentaires

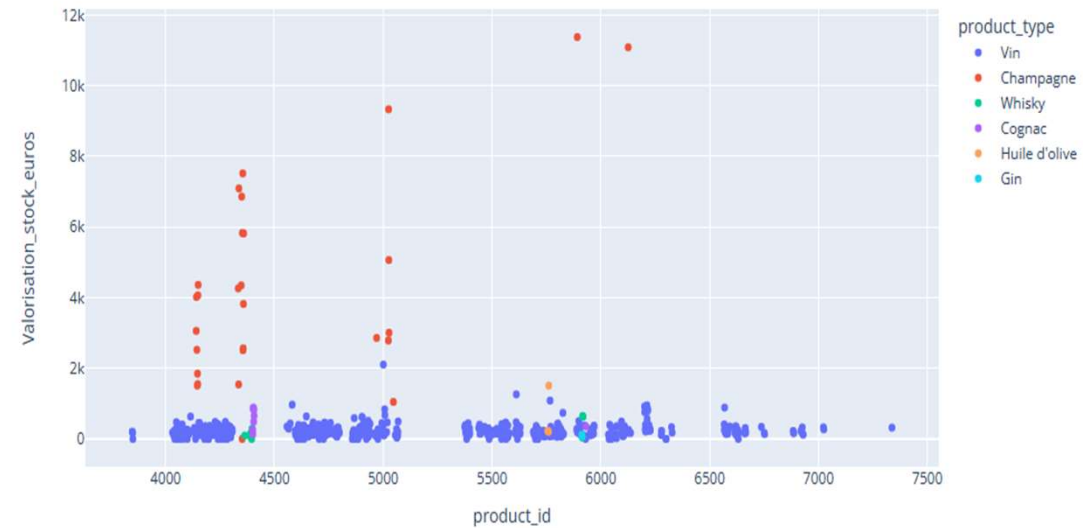
## CA, quantités, stocks, taux de marge et correlations

### Comparaison « Nuage de points » CA en € / Stock en €

CA par type de produit



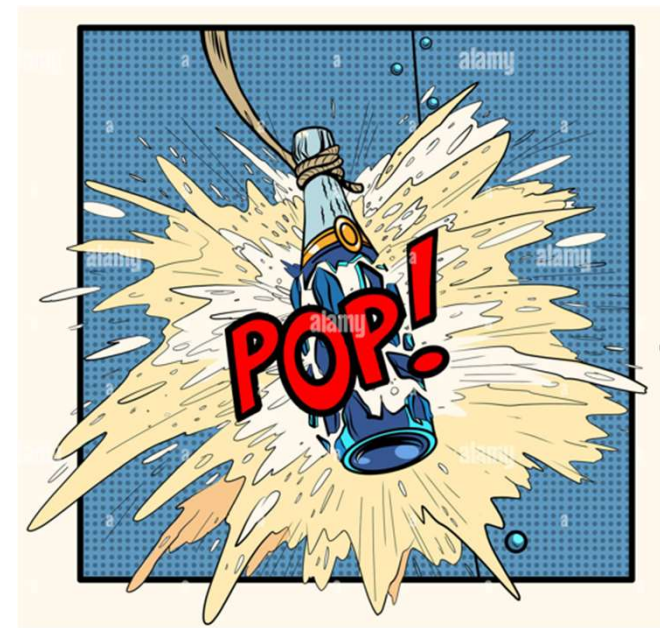
Valorisation du stock par type de produit



## Analyses complémentaires CA, quantités, stocks, taux de marge et correlations

---

- *Point de vigilance sur la famille « Champagne »*
- ❖ *17 références sur les 20 plus mauvaises rotations de stock*
- ❖ *Faible taux de marge à 4,79 %*
- ❖ *1 seule référence dans le Top 20 quantités vendues*
- ❖ *Point positif : 9 références dans le Top 20 CA*



# Actions pour la suite

---

## ➤ *Recommandations pour l'ERP*

- ❖ *Simplifier et rendre les 3 fichiers cohérents et interopérables*
- ❖ *Une seule clé primaire pour faciliter l'analyse : product\_id*
- ❖ *Paramétrer le logiciel des données pour éviter des erreurs de saisie (ex : « prix de vente » toujours positif, valeurs vides interdites pour « product\_id », menu déroulant pour « onsale\_web »...)*
- ❖ *Ajouter une variable : nombre de clients*
- ❖ *Mettre en place des analyses comparatives pour mesurer l'évolution N/N-1*
- ❖ *Détecter et corriger les erreurs, les doublons et les incohérences (importance de l'année...)*
- ❖ *Intégration des fichiers Excel avec les données mises à jour dans Jupyter Notebook pour un reporting rapide et sans effort, en un seul clic...*

# Actions pour la suite

---



*Exemple de l'intérêt pour la détection et le contrôle...*

## Champagne Egly-Ouriet Grand Cru Blanc de Noirs

Prix d'achat HT : 77,48 €

Prix de vente HT : 10,54 €

Perte par UV : 66,94 €

Taux de marge : - 635 % !!!

Stock : 97

Perte potentielle : (66,94 € \* 97 )

Soit **6 493,18 €...**



❖ *Sanction pénale et financière pour vente à perte en cas de contrôle par les fraudes...*

# Point sur les compétences apprises

---

*Qu'est-ce qui s'est bien passé pour vous dans ce travail de nettoyage ?*  
*Supprimer des données incohérentes puis vérifier l'exactitude.*

*Qu'est-ce que vous avez trouvé le plus difficile ?*  
*Comprendre l'interopérabilité des 3 fichiers et détecter les erreurs.*

*Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ?*

*Maîtriser le logiciel Python*

*Maîtriser l'analyse statistique*

**MERCI POUR VOTRE ATTENTION**

