



Optimiser la gestion des données d'une boutique

Parcours Data Analyste
MARCHAND Emmanuel – 05/2024

Analyses exploratoires des données

Données ERP

- Aucun doublon sur la clé article
- Redressage des niveaux de stock nul
- Redressage des statuts de stock
- Correction des stocks négatifs
- Correction des prix négatifs

Données Web

- Suppression de 22 colonnes inutiles à l'analyse
- Suppression de SKU hors règle de codification
- Suppression des lignes sans SKU
- Filtrage des données web sur les post-type équivalent à Product.

Données de Liaison

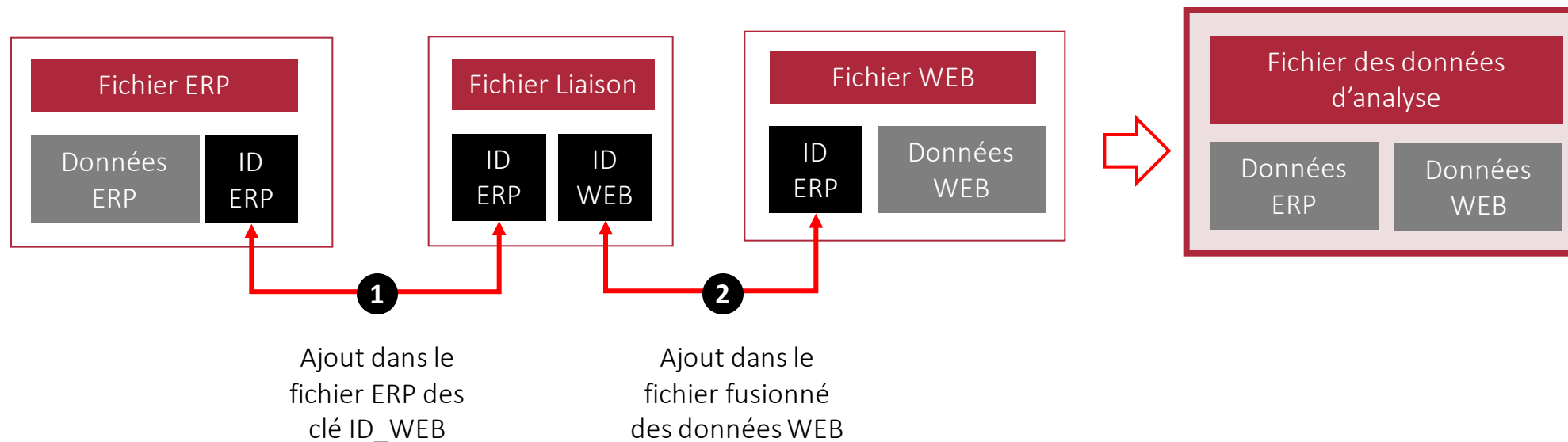
- Attention, présence de doublons sur la clé ID_Web
- Détection de 91 lignes sans ID_Web



REMARQUES – PIEGES :

- Difficulté à identifier l'usage de certaines variables (Onsale_web)
- La recherche du nombre de valeurs nul dans chaque colonne permet d'affiner les filtrages

Fusion des données – en deux étapes



REMARQUES – PIEGES :

- Le choix du type de jointure est important
- L'identification des clés manquantes ou des informations manquantes permet de mieux filtrer

Fusion des données – vue détaillée

Clé de fusion

Données ERP				Données de Liaison			
#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64	0	id_web	734 non-null	object
1	onsale_web	825 non-null	int64	1	product_id	825 non-null	int64
2	price	825 non-null	float64				
3	stock_quantity	825 non-null	int64				
4	purchase_price	825 non-null	float64				
5	stock_status	825 non-null	object				

1

Données Merge				Données Web			
#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64	0	sku	712 non-null	object
1	onsale_web	825 non-null	int64	1	total_sales	712 non-null	float64
2	price	825 non-null	float64	2	tax_status	712 non-null	object
3	stock_quantity	825 non-null	int64	3	product_type	712 non-null	object
4	purchase_price	825 non-null	float64	4	post_title	712 non-null	object
5	stock_status	825 non-null	object	5	post_name	712 non-null	object
6	id_web	734 non-null	object	6	post_type	712 non-null	object

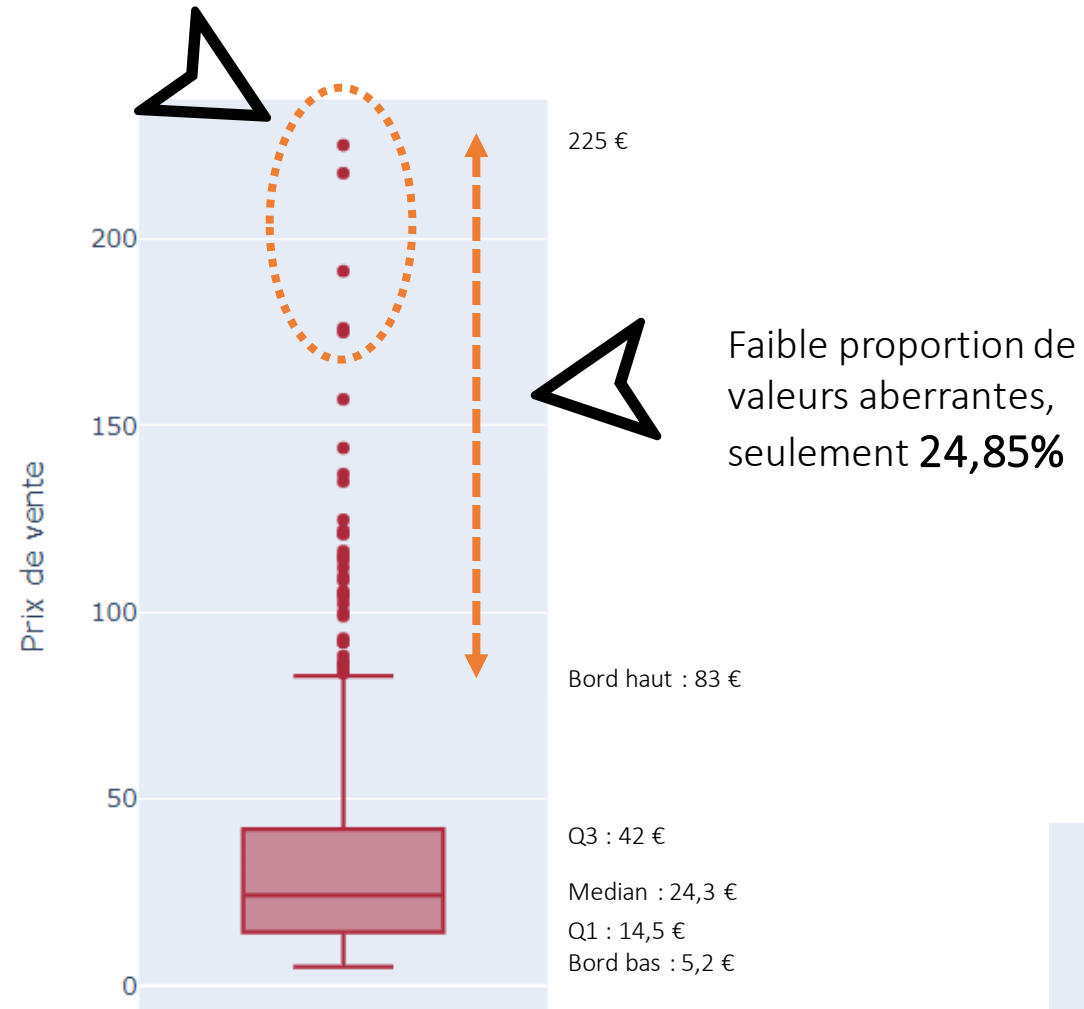
Clé de fusion

2

Données Analyse			
#	Column	Non-Null Count	Dtype
0	product_id	825 non-null	int64
1	onsale_web	825 non-null	int64
2	price	825 non-null	float64
3	stock_quantity	825 non-null	int64
4	purchase_price	825 non-null	float64
5	stock_status	825 non-null	object
6	id_web	734 non-null	object
7	total_sales	712 non-null	float64
8	tax_status	712 non-null	object
9	product_type	712 non-null	object
10	post_title	712 non-null	object
11	post_name	712 non-null	object
12	post_type	712 non-null	object

Analyses univariées du prix

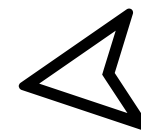
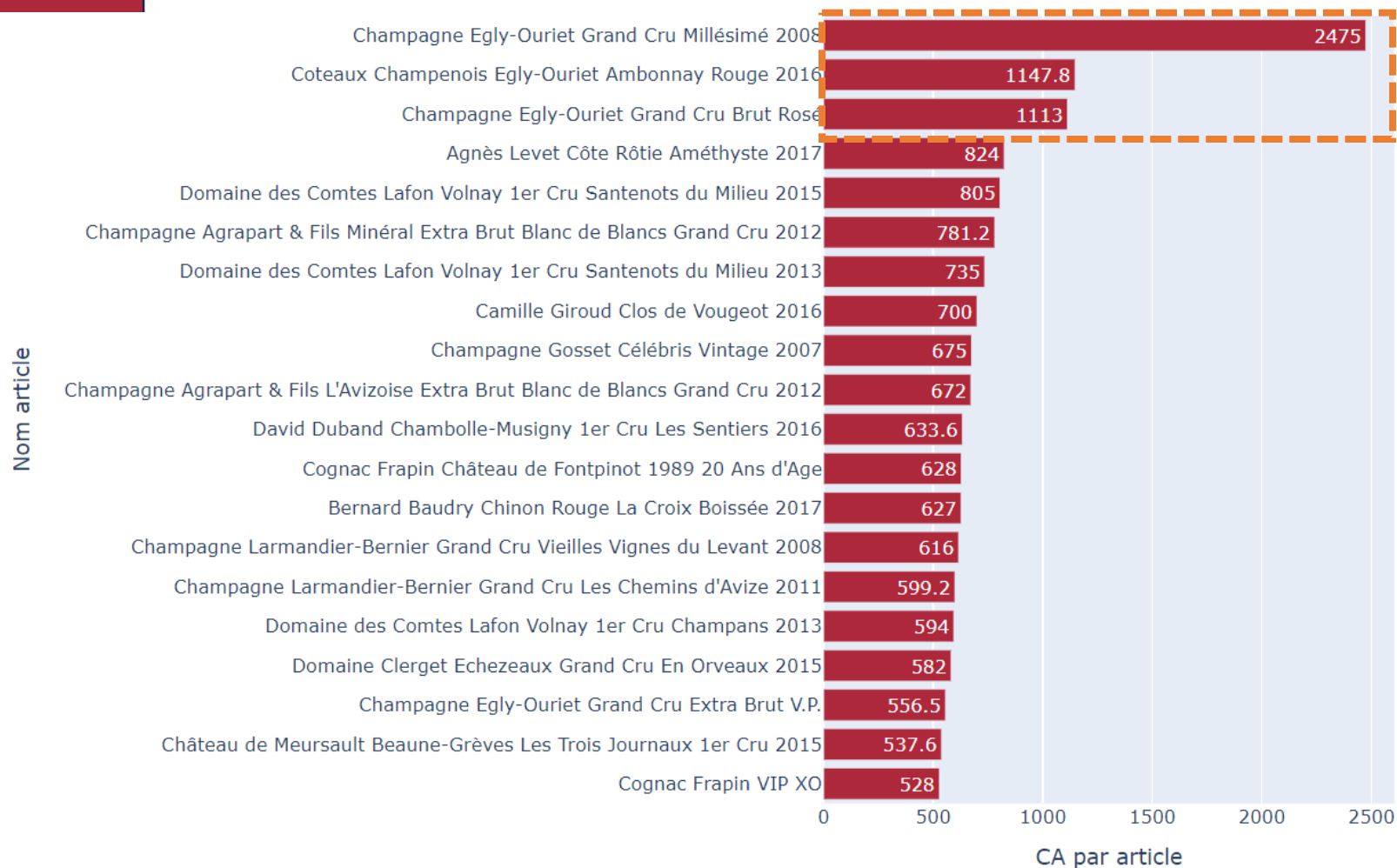
Très peu d'articles
avec des valeurs
extrêmes, seulement
17 articles



Prix moyen : 32,27 €

Ecart type : 26,60 €

Analyses univariées du Chiffre d'Affaire

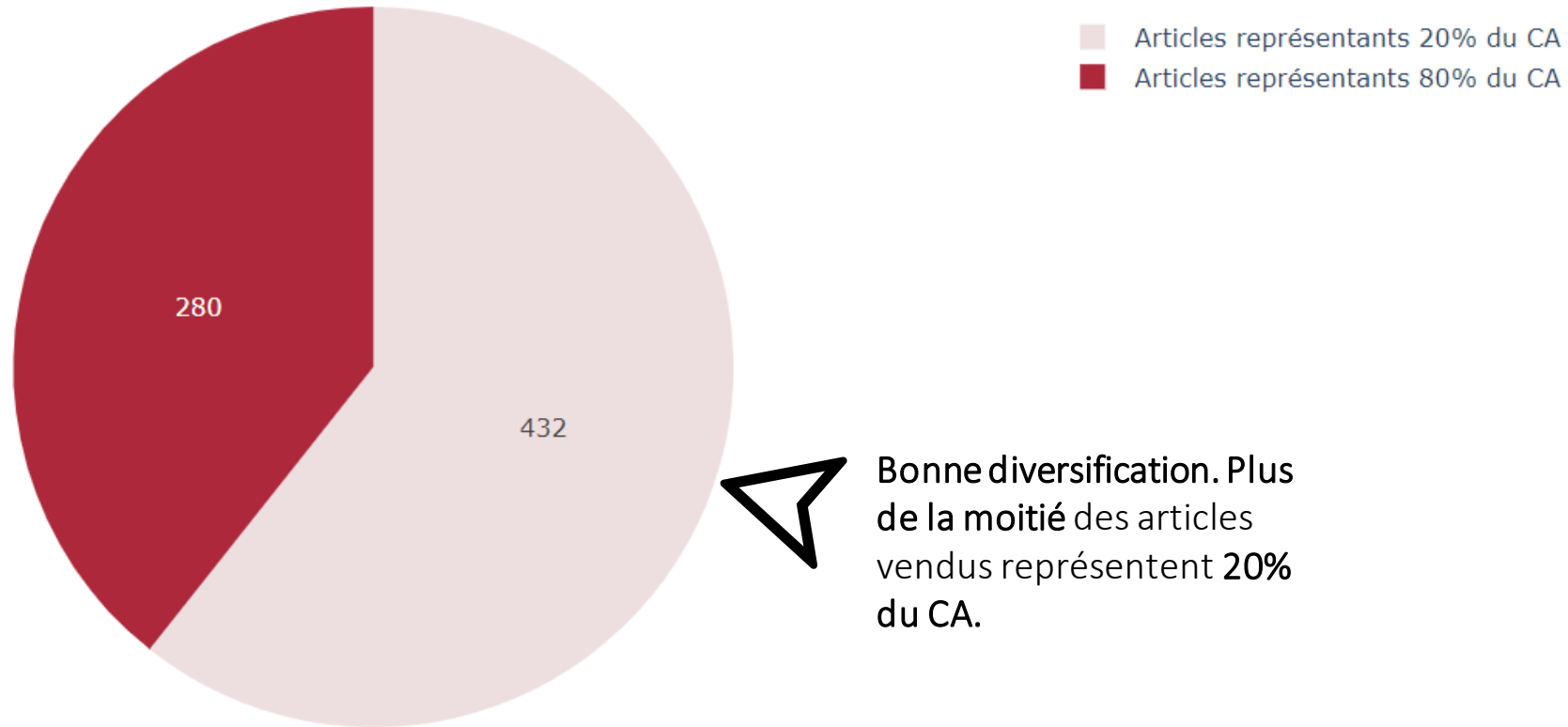


Répartition homogène des CA, hormis le TOP3 qui se distingue nettement.

Le CA total du site web est **143 285 €**

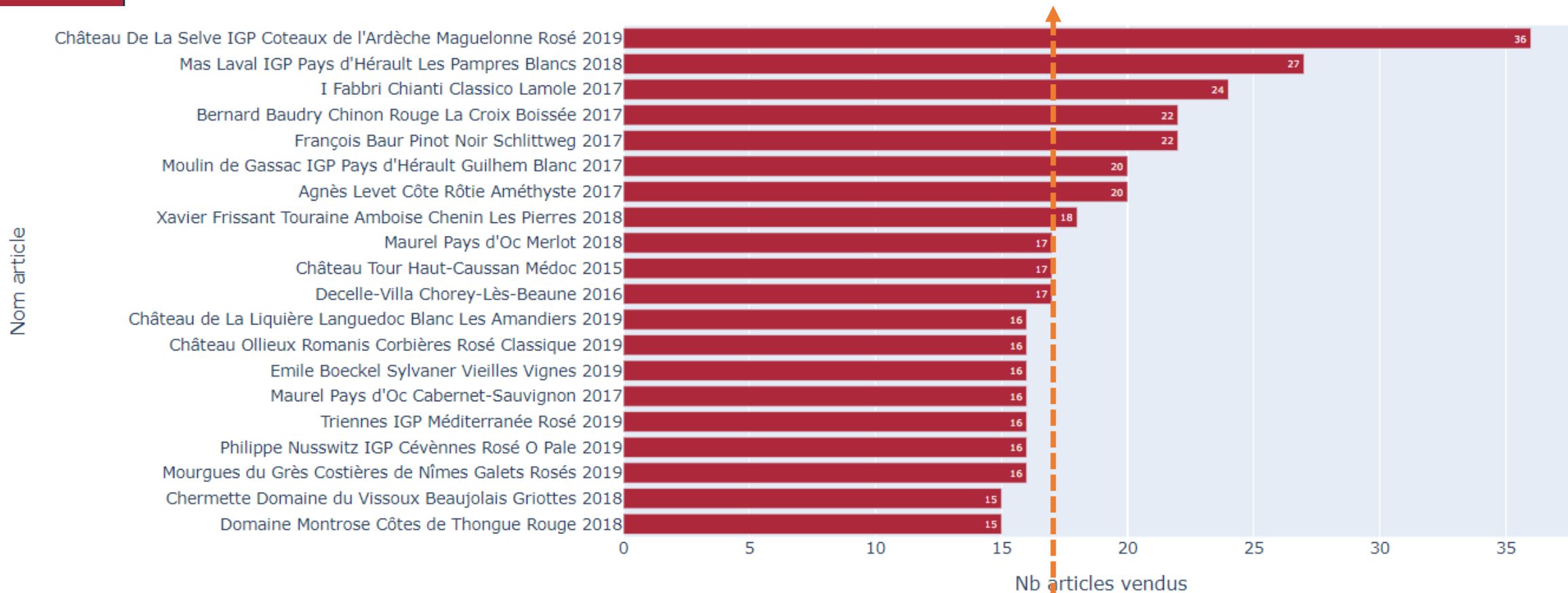
TOP 20 des articles avec le plus gros CA

Analyses univariées du Chiffre d’Affaire



Répartition du nombre d'article représentant
80-20% du CA

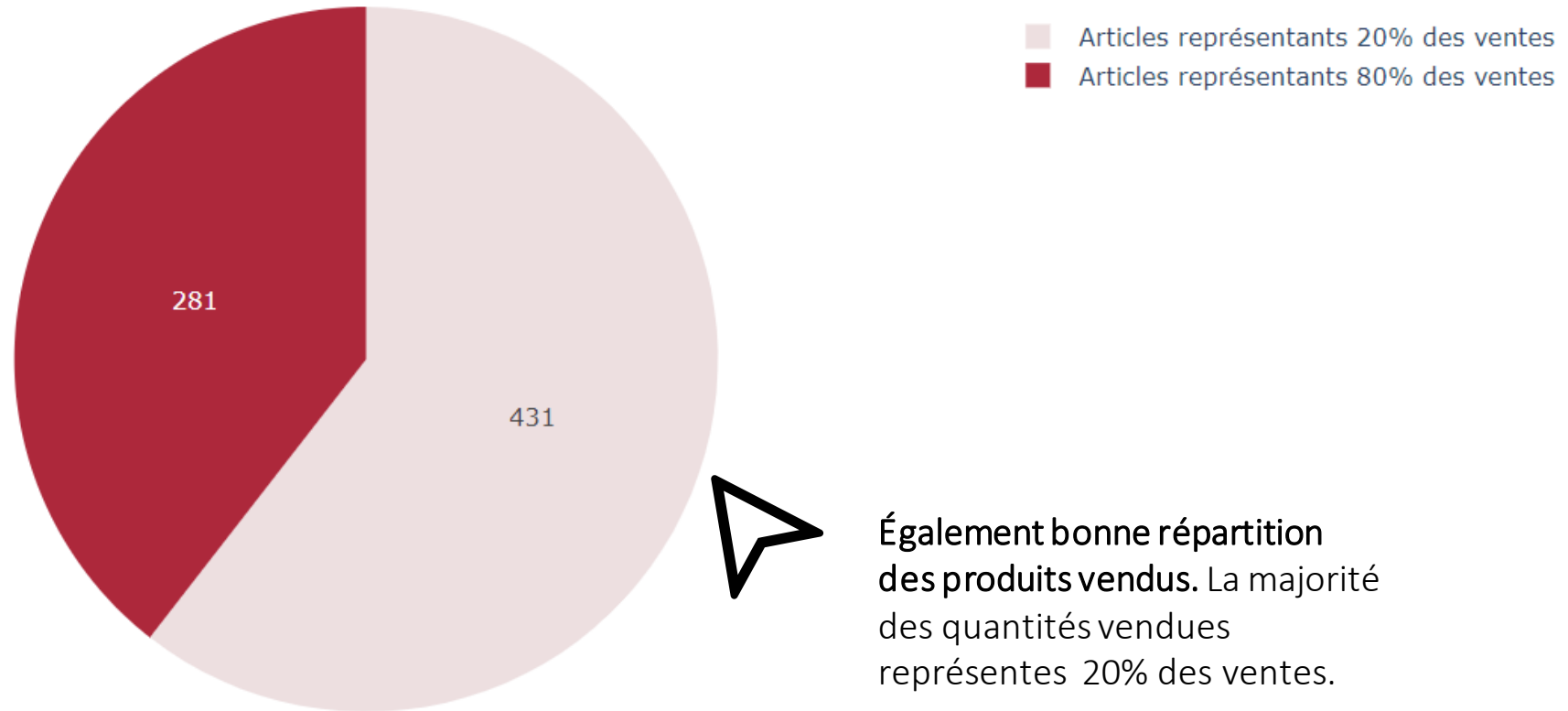
Analyses univariées des ventes en Quantités



TOP 20 des articles les plus vendus

Répartition des quantités les plus vendues autour d'un seuil de 18 articles.

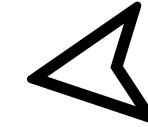
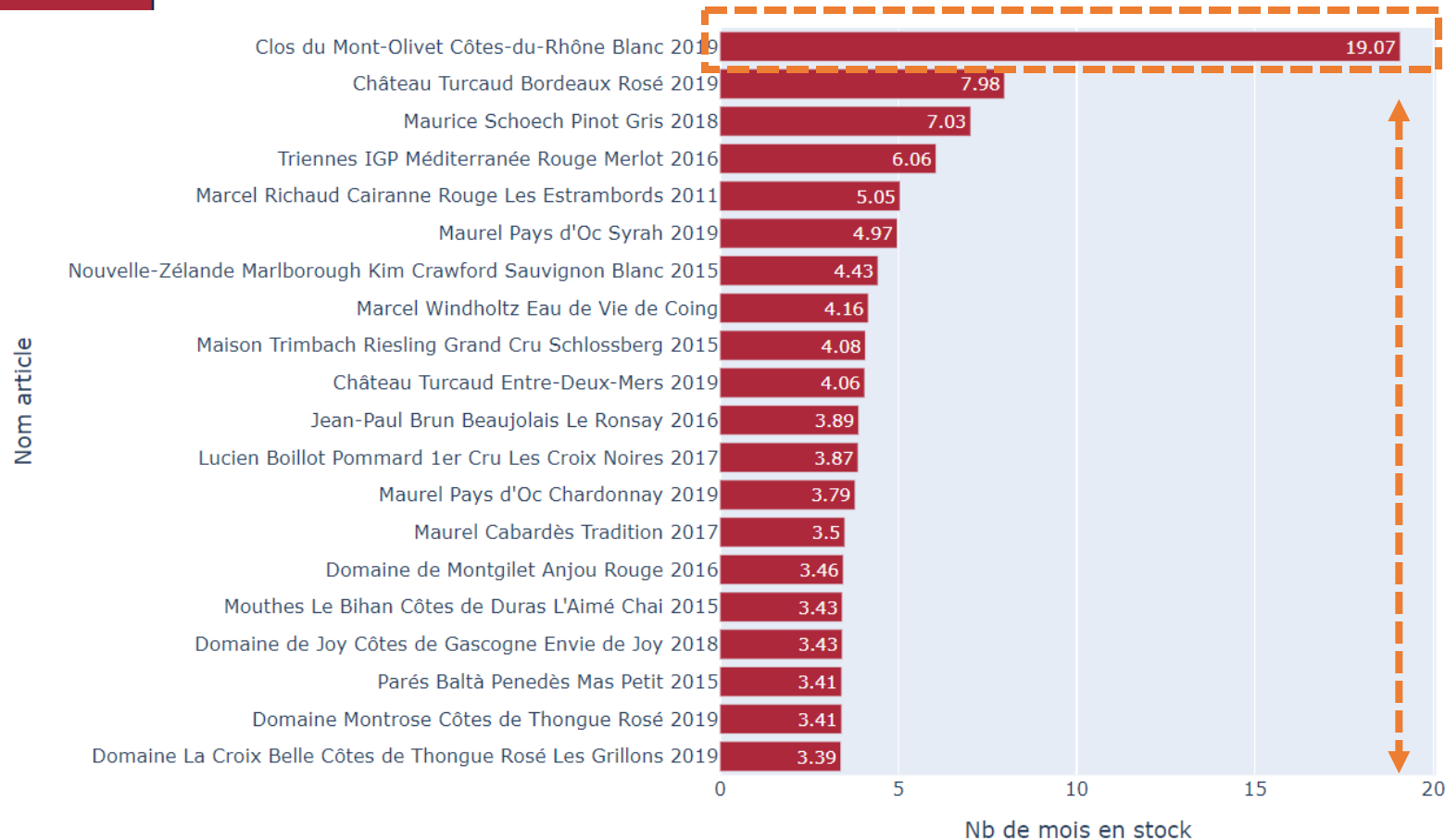
Analyses univariées des ventes en Quantités



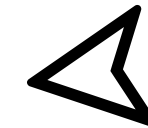
Également bonne répartition des produits vendus. La majorité des quantités vendues représente 20% des ventes.

Répartition du nombre d'article représentant 80-20% du nb de ventes

Analyses univariées des Stocks



Pas d'alerte, cela ne concerne que 3 articles en stock.



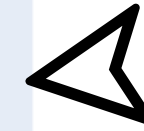
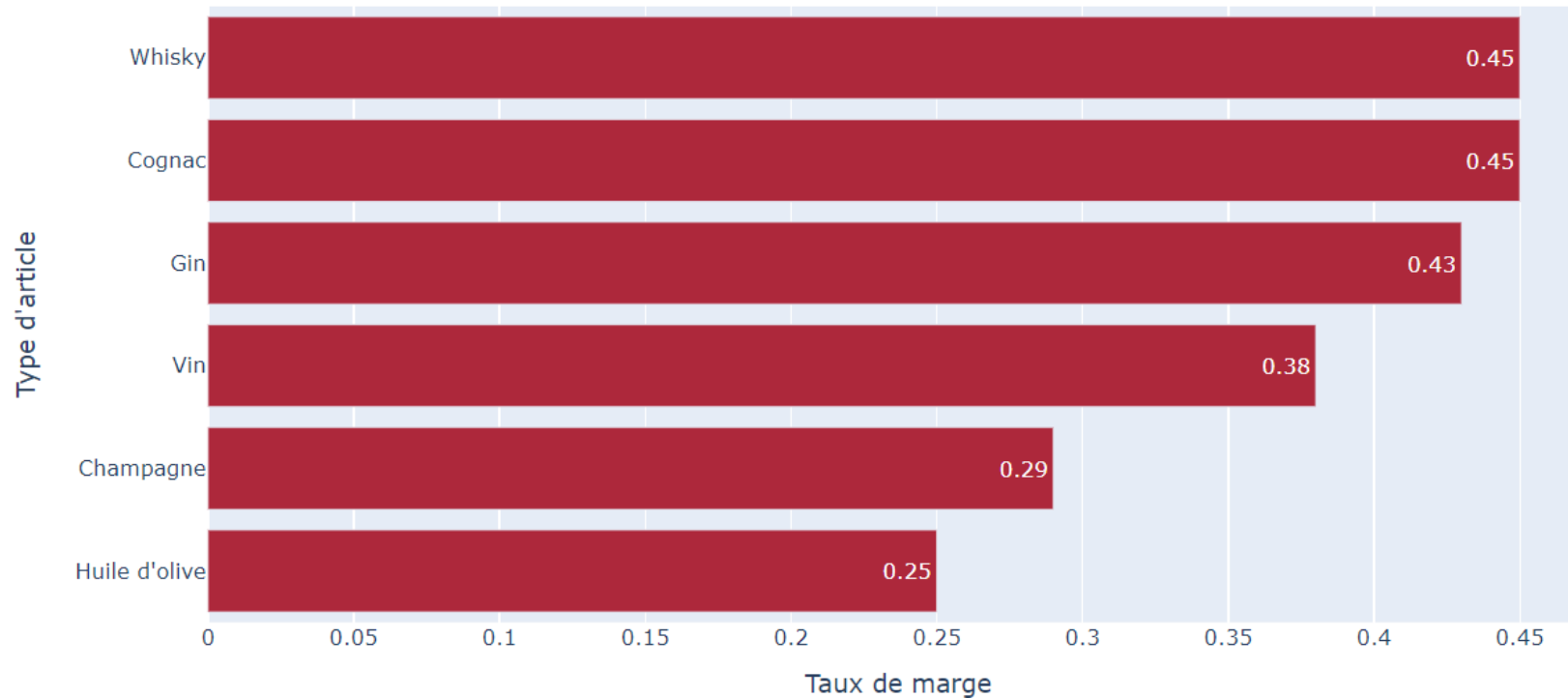
Analyse nécessaire entre les types haut de gamme ou de consommation courante pour définir les bons KPI de stock (rotation et coûts de stockage).

Nb d'articles en stocks : 17 833

Valorisation des stocks : 298 699€

FLOP 20 des articles les plus en stocks
(en nombre de mois en stock)

Analyses univariées du Taux de Marge



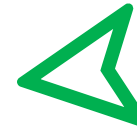
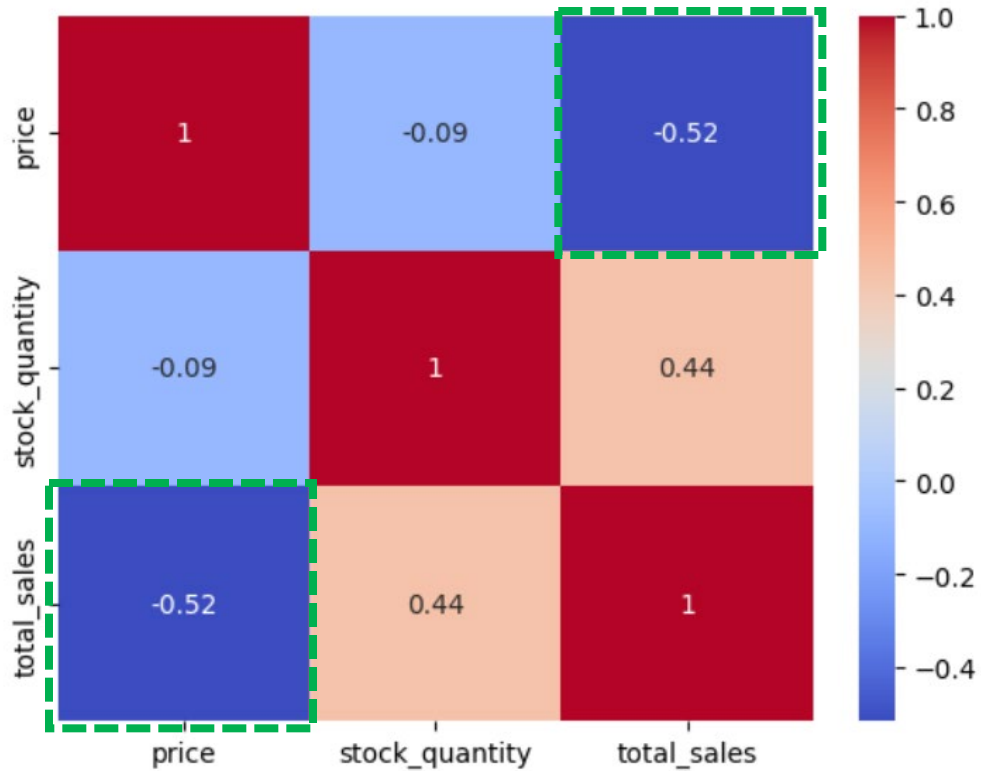
Taux de marge très faible pour l'ensemble des articles (*).

Etude à prévoir sur les prix d'achat et les prix de vente.

Taux de marge par type d'article

(*) Dans les secteurs de la distribution ou les services, un taux de marge de 10 à 15 % peut être considéré comme acceptable.

Analyse multivariée



Forte corrélation entre le prix et les ventes.

Les niveaux de stock influent un peu sur les ventes et n'ont aucuns liens sur les prix.

Heatmap de corrélation entre les prix, les stocks et les ventes

Actions pour la suite

1

Automatiser la production de cette étude, pour cela il faudra :

- ❌ **Supprimer les doublons** d'articles dans les extractions de l'ERP.
- ❌ Travailler sur **la qualité** des informations **de stock**
- ❌ Faire les **évolutions nécessaires** pour empêcher la saisie de données négatives dans le système (stock, prix)
- ❌ **Alléger** les fichiers d'exportation des données Web
- ❌ **Verrouiller le système** pour contraindre le respect des règles de codification article
- ❌ Travailler sur **la qualité** des données **de liaison** et les rafraichir le plus souvent possible.

2

Capitaliser sur cette étude pour préparer l'avenir :

- ❌ **Définir des** indicateurs clé de performance (KPI) pour les sujets étudiés :
 - Les prix
 - Le CA
 - Les ventes
 - Les stocks
 - Le taux de marge
 - Les corrélations
- ❌ Définir et construire un **tableau de bord mensuel**.

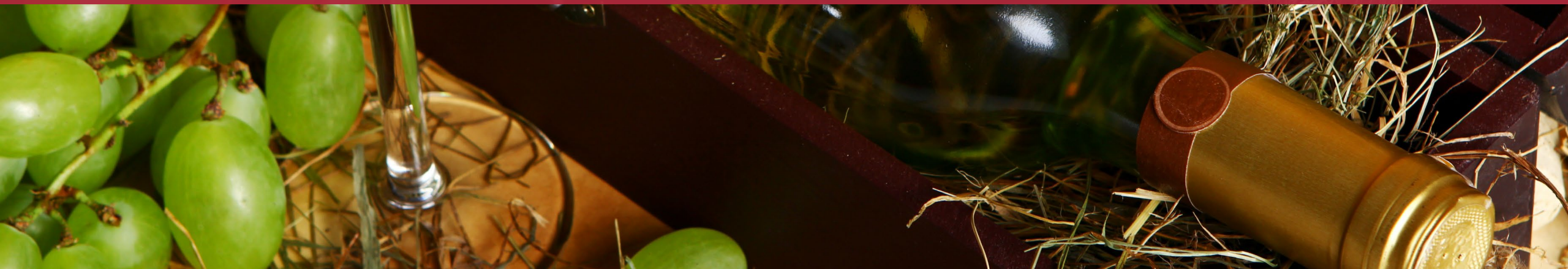


Point sur les compétences acquises

- Il est difficile de devoir nettoyer les données **sans supprimer les informations primordiales** => en situation, réunions avec les sachants métiers de la société.
- Acquisition de nouvelles compétences **sur l'analyse statistique** : Z-score, interquartile, Pareto.
- Acquisition de basiques sur **le contrôle de gestion** : utilisation du Pareto pour le CA et les ventes et des sommes cumulatives, le calcul de rotation des stocks, le calcul de marge commerciale et de taux de marge.
- A voir : comment calculer **le taux de rentabilité** d'un produit.
- Le **diagramme à moustache** est un très bon outil d'analyse
- Premiers pas avec **plotly express** : temps passé pour customiser les graphiques.
- La production de données raffinées n'est pas la finalité, **produire des informations** pour la société est le plus difficile et sera un travail d'amélioration continue.



MERCI



Ressources



ATTRIBUTION images :

https://fr.freepik.com/photos-gratuite/fond-vignoble-culture-du-raisin-paysage-agricole_136712307.htm#fromView=search&page=1&position=0&uuid=06a42bee-a0cd-4e62-b79f-67add9fdf4da>Image de frimufilms sur Freepik

https://fr.freepik.com/photos-gratuite/belle-composition-vin_7907039.htm#fromView=search&page=1&position=24&uuid=06a42bee-a0cd-4e62-b79f-67add9fdf4da>Image de Racool_studio sur Freepik

Couleur rouge : #AD283B