

OBJECTIFS DE LA RÉUNION :

RÉSULTATS DES VENTES EN LIGNE ANALYSE DE LA VARIABLE PRIX

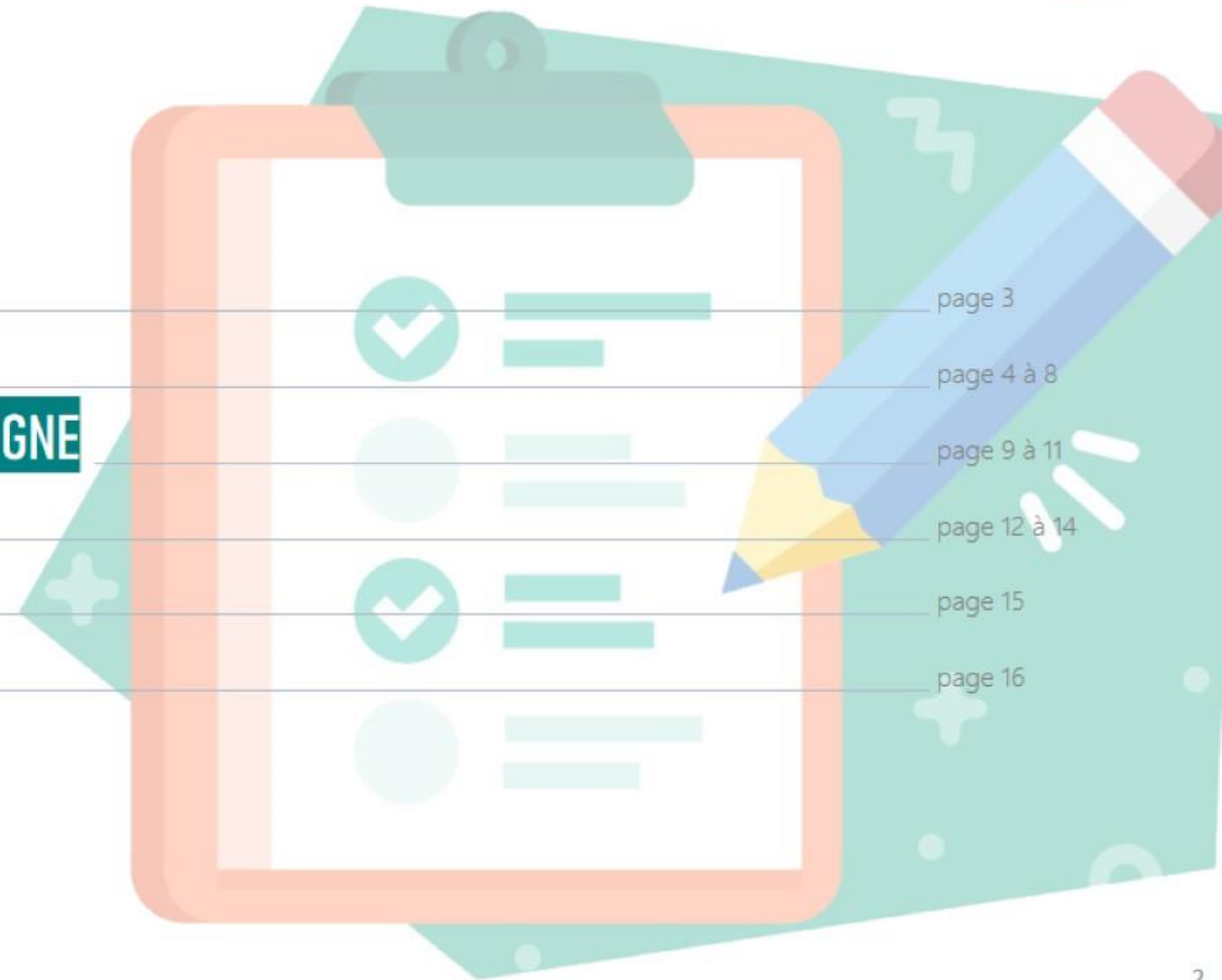
Cruder Léa
Data Analyst



SOMMAIRE



1. CONTEXTE (PROBLÉMATIQUE)
2. MÉTHODOLOGIE
3. RÉSULTATS DU CHIFFRE D'AFFAIRE EN LIGNE
4. ANALYSE DE LA VARIABLE PRIX
5. INCOHÉRENCES DANS LES DONNÉES
6. CONCLUSION



page 3

page 4 à 8

page 9 à 11

page 12 à 14

page 15

page 16

1 - CONTEXTE



ACTUELLEMENT, LES OUTILS DE GESTION NE SONT PAS RELIÉS (ERP ET CMS)

POSE 3 GROSSES PROBLÉMATIQUES :



GESTION DES STOCKS SUR L'ERP
VISIBILITÉ RÉDUITE POUR ANALYSER LES VENTES EN LIGNE
ERREURS POTENTIELLES DE SAISIE, NOTAMMENT SUR LES PRIX

DANS L'ATTENTE D'UNE SOLUTION CENTRALISÉE, NOUS FERONS UN RAPPROCHEMENT MANUEL ENTRE LES 2 POUR
POUVOIR ANALYSER LE CHIFFRE D'AFFAIRE RÉALISÉ SUR LE SITE, PUIS NOUS CHERCHERONS COMMENT DÉCELER
LES ERREURS POTENTIELLES DE PRIX AFIN DE LES ANTICIPER AU MIEUX.

2.1 - MÉTHODOLOGIE



2.2 - ANALYSE EXPLORATOIRE ET NETTOYAGE DE L'EXTRAIT ERP



fichierERP

	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
...
820	7203	0	45.0	30	instock
821	7204	0	45.0	9	instock
822	7247	1	54.8	23	instock
823	7329	0	26.5	14	instock
824	7338	1	16.3	45	instock

825 rows × 5 columns

CORRESPOND AUX VENTES SUR LE SITE :
1 POUR OUI, 0 POUR NON -> FILTRAGE

CONTIENT 2 VALEURS NÉGATIVES
-> SUPPRESSION

- FILTRAGE DANS UNE COPIE
- VÉRIFICATION DE L'UNICITÉ DE LA CLÉ PRODUCT_ID
- PAS DE DOUBLONS
- PAS DE VALEURS NULLES
- SUPPRESSION DES COLONNES INUTILES / RENOMMAGE

Affichage

fichierERPsite

	id_produit	prix
0	3847	24.2
1	3849	34.3
2	3850	20.8
3	4032	14.1
4	4039	46.0
...
787	6930	8.4
792	7023	27.5
793	7025	69.0
822	7247	54.8
824	7338	16.3

717 rows × 2 columns

2.3 - ANALYSE EXPLORATOIRE ET NETTOYAGE DE L'EXTRAIT DU SITE



Affichage des infos du dataset

```
fichierWeb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1513 entries, 0 to 1512
Data columns (total 28 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sku                   1428 non-null   object
1   virtual              1513 non-null   int64
2   downloadable         1513 non-null   int64
3   rating_count        1513 non-null   int64
4   average_rating      1430 non-null   float64
5   total_sales         1430 non-null   float64
6   tax_status          716 non-null    object
7   tax_class            0 non-null      float64
8   post_author         1430 non-null   float64
9   post_date           1430 non-null   datetime64[ns]
10  post_date_gmt       1430 non-null   datetime64[ns]
11  post_content        0 non-null      float64
12  post_title          1430 non-null   object
13  post_excerpt        716 non-null    object
14  post_status         1430 non-null   object
15  comment_status      1430 non-null   object
16  ping_status         1430 non-null   object
17  post_password       0 non-null      float64
18  post_name           1430 non-null   object
19  post_modified       1430 non-null   datetime64[ns]
20  post_modified_gmt   1430 non-null   datetime64[ns]
21  post_content_filtered 0 non-null      float64
22  post_parent         1430 non-null   float64
23  guid                1430 non-null   object
24  menu_order          1430 non-null   float64
25  post_type           1430 non-null   object
26  post_mime_type      714 non-null    object
27  comment_count       1430 non-null   float64
dtypes: datetime64[ns](4), float64(10), int64(3), object(11)
memory usage: 331.1+ KB
```

- SUPPRESSION DES DOUBLONS
- SUPPRESSION DES VALEURS NULLES
- VÉRIFICATION DE L'UNICITÉ DE LA CLÉ SKU
- SUPPRESSION DES COLONNES INUTILES / RENOMMAGE

Affichage du dataset

```
fichierWeb
```

	SKU	total_ventes	nom_produit
0	16004	5.0	Château du Couvent Pomerol 2017
2	15075	3.0	Parés Baltà Penedès Indigena 2017
3	16209	6.0	Maurel Cabardès Tradition 2017
4	15763	1.0	Domaine de la Monardière Vacqueyras Vieilles V...
5	13895	0.0	Château Saransot-Dupré Bordeaux Blanc 2016
...
715	15575	6.0	Château Plaisance Fronton Rouge 2017
716	15857	0.0	Domaine de Montbourgeau L'Etoile En Banode 2016
717	16011	0.0	Nouvelle-Zélande Marlborough Momo Pinot Noir 2016
718	15180	1.0	Camille Giroud Santenay Rouge 2016
719	1662	87.0	Champagne Gosset Grand Rosé

714 rows x 3 columns

2.4 - ANALYSE EXPLORATOIRE ET NETTOYAGE DU FICHIER DE LIAISON



CORRESPOND AU SKU
CONTIENT 91 VALEURS MANQUANTES

fichierLiaison

	product_id	id_web
0	3847	15298
1	3849	15296
2	3850	15300
3	4032	19814
4	4039	19815
...
820	7203	NaN
821	7204	NaN
822	7247	13127-1
823	7329	14680-1
824	7338	16230

825 rows × 2 columns

- RENOMMAGE DES COLONNES
- SUPPRESSION DES VALEURS MANQUANTES SUR SKU (ID_WEB)
- VÉRIFICATION DE L'UNICITÉ DES CLÉS

Affichage du dataset

fichierLiaison

	id_produit	SKU
0	3847	15298
1	3849	15296
2	3850	15300
3	4032	19814
4	4039	19815
...
792	7023	15891
793	7025	15887
822	7247	13127-1
823	7329	14680-1
824	7338	16230

734 rows × 2 columns

2.5 - CROISEMENT DES DONNÉES (JOINTURES)



**JOINTURE INTERNE
ENTRE LE FICHIER ERP
ET LE FICHIER LIAISON**

	id_produit	prix	SKU
0	3847	24.2	15298
1	3849	34.3	15296
2	3850	20.8	15300
3	4032	14.1	19814
4	4039	46.0	19815
...
709	6930	8.4	16135
710	7023	27.5	15891
711	7025	69.0	15887
712	7247	54.8	13127-1
713	7338	16.3	16230

714 rows × 3 columns

**PUIS 2^E JOINTURE INTERNE
AVEC LE FICHIER WEB**

	id_produit	prix	SKU	total_ventes	nom_produit
0	3847	24.2	15298	6.0	Pierre Jean Villa Saint-Joseph Préface 2018
1	3849	34.3	15296	0.0	Pierre Jean Villa Saint-Joseph Rouge Tildé 2017
2	3850	20.8	15300	0.0	Pierre Jean Villa Crozes-Hermitage Accroche Co...
3	4032	14.1	19814	3.0	Pierre Jean Villa IGP Collines Rhodaniennes Ga...
4	4039	46.0	19815	0.0	Pierre Jean Villa Côte Rôtie Carmina 2017
...
709	6930	8.4	16135	5.0	Mouthes Le Bihan Côtes de Duras L'Aimé Chai 2015
710	7023	27.5	15891	0.0	Camin Larredya Jurançon Sec La Virada 2018
711	7025	69.0	15887	0.0	Domaine Jamet Côte Rôtie Fructus Voluptas 2018
712	7247	54.8	13127-1	0.0	Clos du Mont-Olivet Châteauneuf-du-Pape 2007
713	7338	16.3	16230	0.0	Domaine Saint-Nicolas Vin de France Blanc Les ...

714 rows × 5 columns

**VÉRIFICATION À CHAQUE
ÉTAPE**

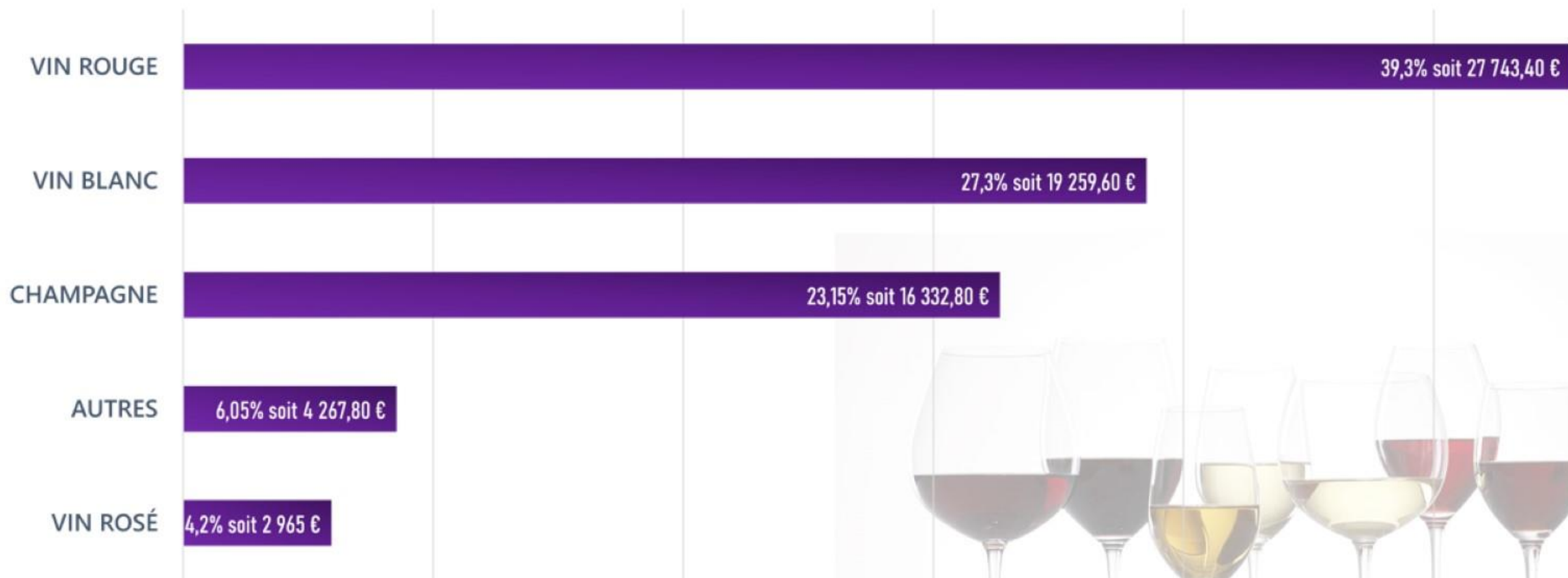
3.1 - CHIFFRE D'AFFAIRE RÉALISÉ EN LIGNE

TOTAL : 70 569 €

NOMBRE D'ARTICLES VENDUS : 2865



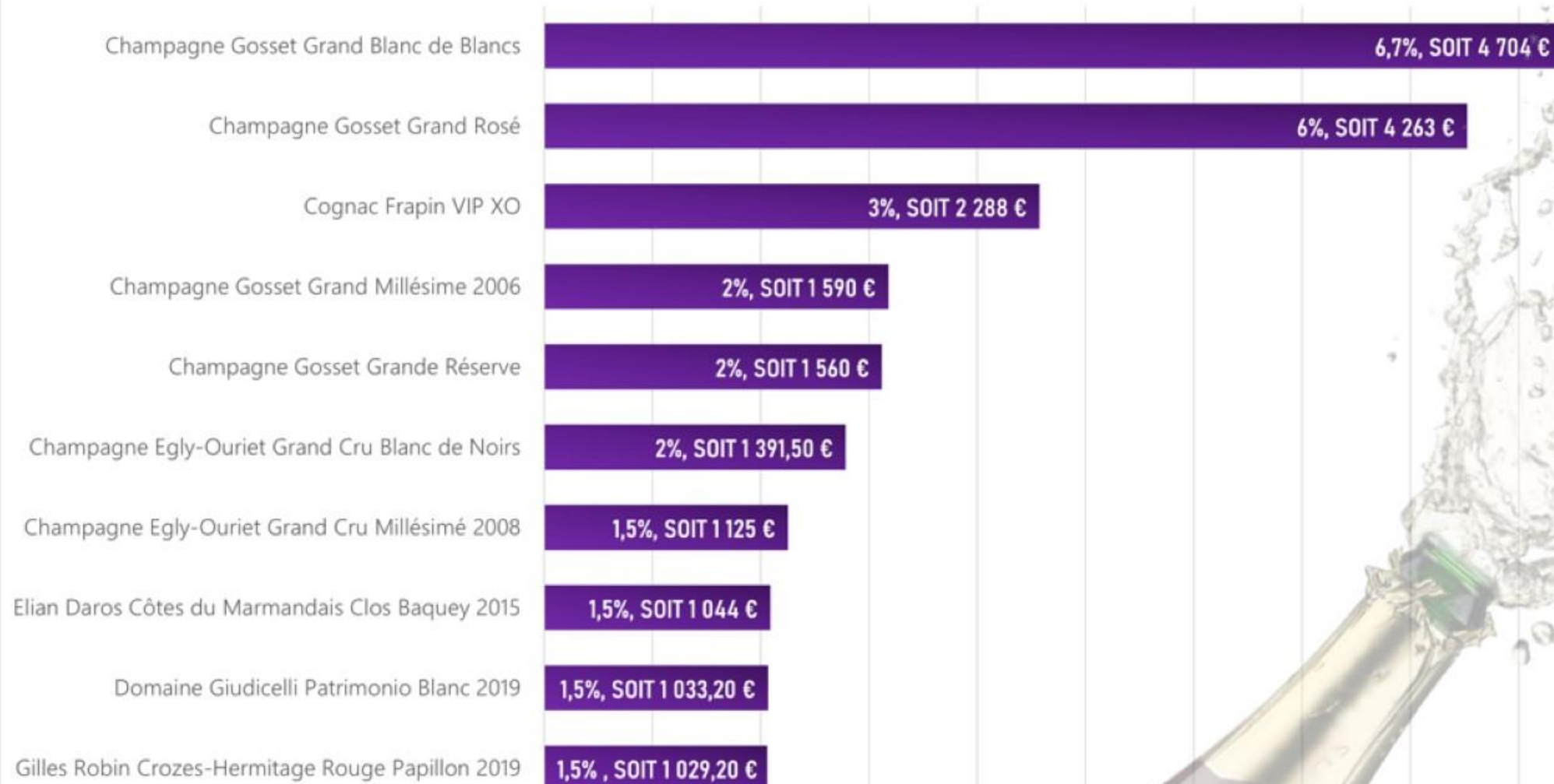
RÉPARTITION DU CHIFFRE D'AFFAIRE



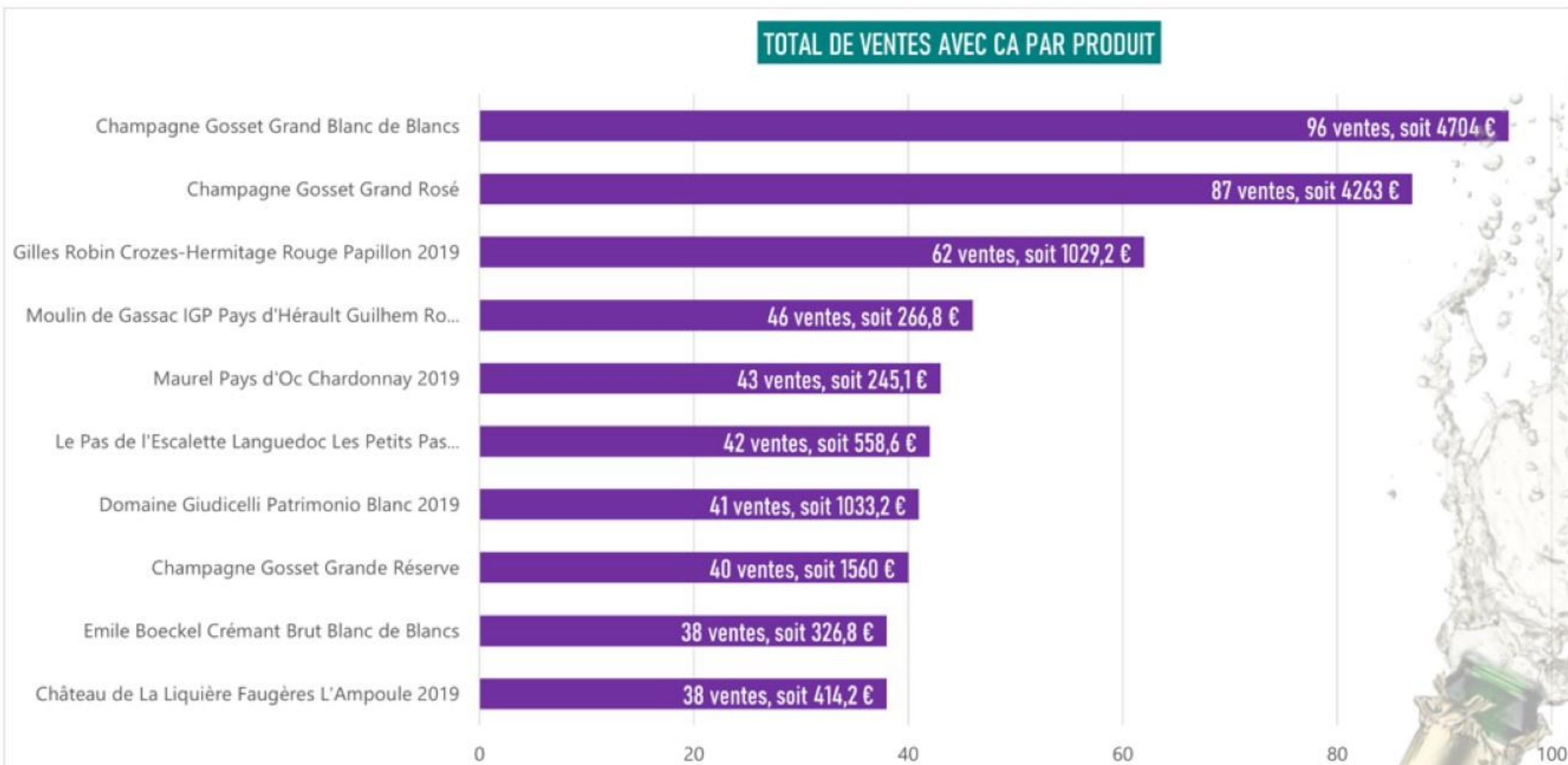
3.2 - TOP 10 DES PRODUITS LES PLUS RENTABLES



TOP 10 DES PRODUITS AVEC LA PART DU CA TOTAL EN %



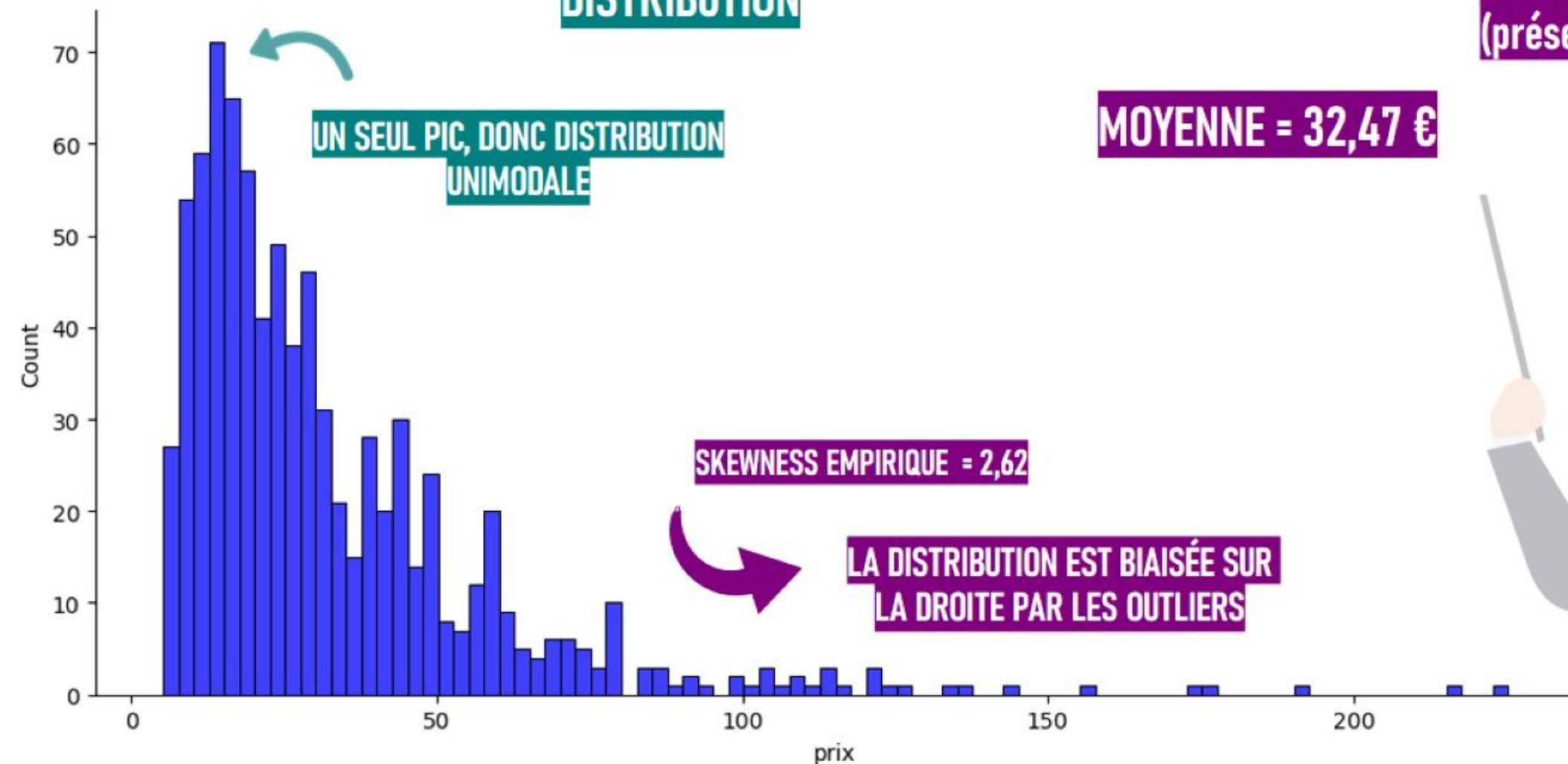
3.3 - TOP 10 DES PRODUITS LES PLUS VENDUS



4.1 - ANALYSE DE LA VARIABLE PRIX : DISTRIBUTION



REPRÉSENTATION GRAPHIQUE DE LA DISTRIBUTION



CLASSE MODALE = 45 €
(présent 15 fois)

MOYENNE = 32,47 €

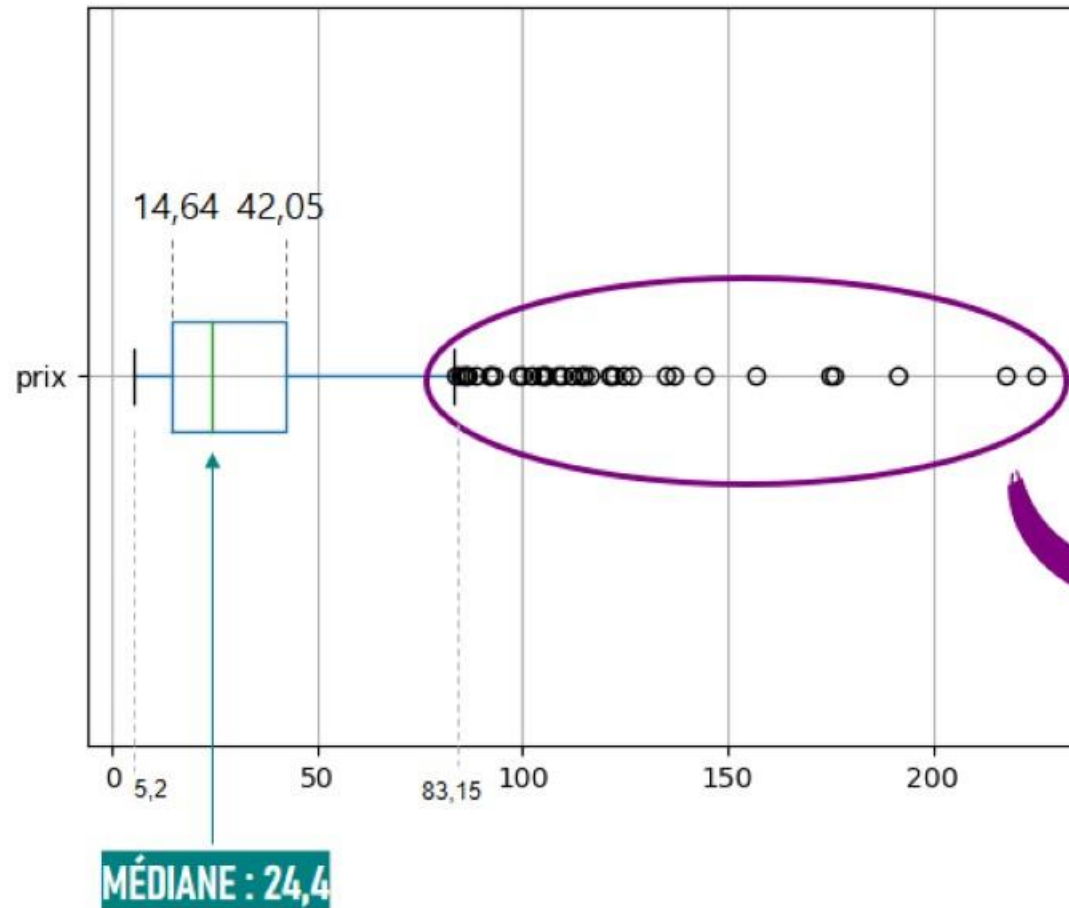
SKEWNESS EMPIRIQUE = 2,62

LA DISTRIBUTION EST BIAISÉE SUR LA DROITE PAR LES OUTLIERS



4.2 - ANALYSE DE LA VARIABLE PRIX : MESURES DE DISPERSION

VISUALISATION DE LA DISTRIBUTION AVEC UNE BOÎTE À MOUSTACHE



IQR = 27,4

BORNE SUP = $Q3 + 1,5 \cdot IQR = 83,15$

50 % DES EFFECTIFS SONT CONTENUS ENTRE
14,64 ET 42,05

37 VALEURS SUPÉRIEURES À LA BORNE SUP.
COMPRIS ENTRE 83,15 € ET 225 €

+ LES 2 VALEURS NÉGATIVES

= 39 ERREURS POTENTIELLES

5 – INCOHÉRENCES DANS LES DONNÉES



DANS LE FICHIER ERP :

2 VALEURS NÉGATIVES DANS LES PRIX

2 VALEURS NÉGATIVES DANS LES STOCKS

DANS LE FICHIER WEB :

82 LIGNES EN DOUBLE

798 DOUBLONS SUR LA CLÉ 'SKU'

PRÉSENCE DE VALEURS NULLES, DONT 85 SUR LA CLÉ

DANS LE FICHIER LIAISON :

91 VALEURS NULLES



CHAQUE INCOHÉRENCE A FAIT
L'OBJET D'UN TRAITEMENT ADAPTÉ
POUR NE PAS FAUSSER LES
RÉSULTATS

6 - CONCLUSION



DANS L'IMMÉDIAT, NOUS N'AVONS PAS LA CERTITUDE QUE CES 39 PRIX SOIENT EFFECTIVEMENT DES ERREURS (SAUF POUR LES PRIX NÉGATIFS), MAIS IL FAUDRAIT LES CONFIRMER.

IL FAUT CEPENDANT GARDER CETTE ÉVENTUALITÉ EN TÊTE POUR LA MISE EN PLACE DE LA SOLUTION CENTRALISÉE.

POUR Y REMÉDIER, IL SERAIT INTÉRESSANT DE METTRE EN PLACE UN SYSTÈME D'ALERTE LORSQU'UN PRIX CONSIDÉRÉ COMME INHABITUEL SERA SAISI, PAR EXEMPLE UN PRIX NÉGATIF OU SUPÉRIEUR À 83,15€.