



Alexandre Duvernay

26 juin 2024

# PROJET 6

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



# Introduction

## Mission

Sur la base des données extraites de l'ERP et du site Web de l'entreprise (accompagnées de leur fichier de liaison), effectuer une **analyse des données de vente de l'entreprise** (prix, stock, CA, marge, etc...).

S'assurer préalablement d'avoir des **données nettoyées et consolidées**.

### *DONNÉES TRANSMISES*



- erp.xlsx
- web.xlsx
- liaison.xlsx

### *ENVIRONNEMENT DE TRAVAIL*



python™



# 1 – Analyses exploratoires des données

## Datasets

Nous disposons de 3 datasets, composés de **tableaux** Excel ayant les caractéristiques suivantes :

DATASET	erp.xlsx	web.xlsx	liaison.xlsx
<b>Structure</b>			
Lignes	825	1513	825
Colonnes	6	29	2
<b>Type de données</b>			
Quantitatives	Oui	Oui	Oui
Qualitatives	Oui	Oui	Oui
Dates	Non	Oui	Non



# 1 – Analyses exploratoires des données

## Traitements réalisés

Sur l'ensemble des données, **les traitements suivants** ont été réalisés :



- ❖ **Recherche de doublons**, de valeurs incohérentes, non renseignées ou nulles.
  - Suppression des doublons
  - Correction de valeurs erronées ou incohérentes (stock\_status, price, sku)
  
- ❖ **Recherche des colonnes** (ou variables) utiles à l'analyse (et conformité au RGPD)
  - Suppression des colonnes inutiles



# 1 – Analyses exploratoires des données

## Remarques



Il est important de **comprendre les données** que l'on traite.

En effet les valeurs incohérentes détectées peuvent être corrigées de différentes manières selon l'analyse que l'on en fait.

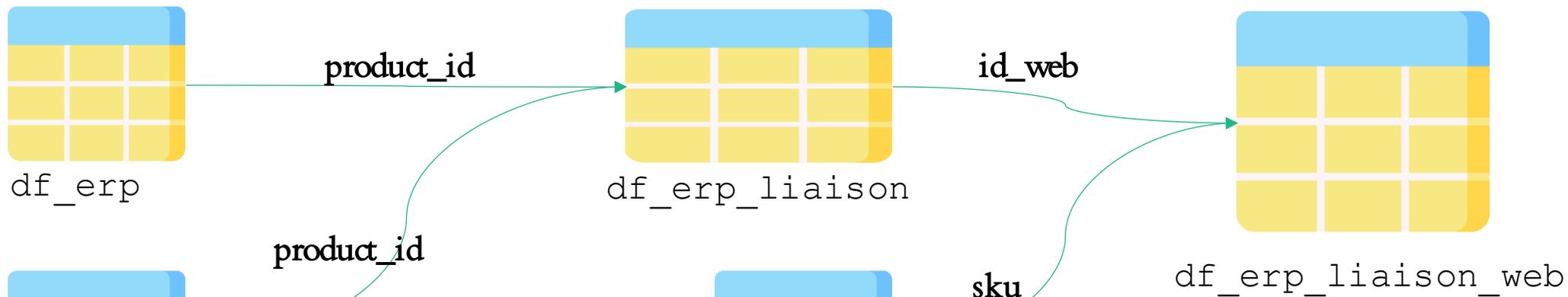
Des **valeurs négatives peuvent être le fruit d'une erreur humaine** à la saisie (price) ou des valeurs légitimes (stock).

L'action à entreprendre pour les nettoyer ne sera évidemment pas la même selon le cas de figure.



# 2 – Fusion ou consolidation des données

Choix des attributs



```
df_erp_liaison =  
pd.merge(df_erp,  
df_liaison)  
# jointure naturelle
```

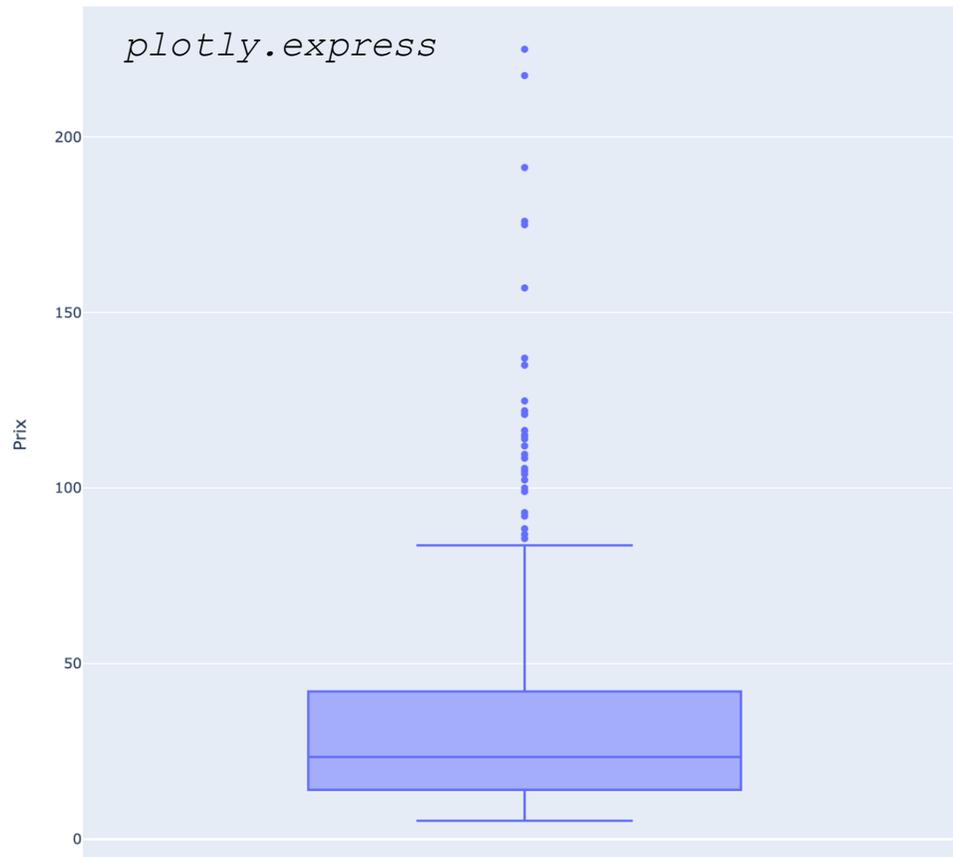
```
df_erp_liaison_web =  
pd.merge(df_web, df_erp_liaison,  
left_on='sku', right_on='id_web',  
how='inner') # jonction interne  
pour écarter les valeurs nulles  
ou sans correspondance.
```



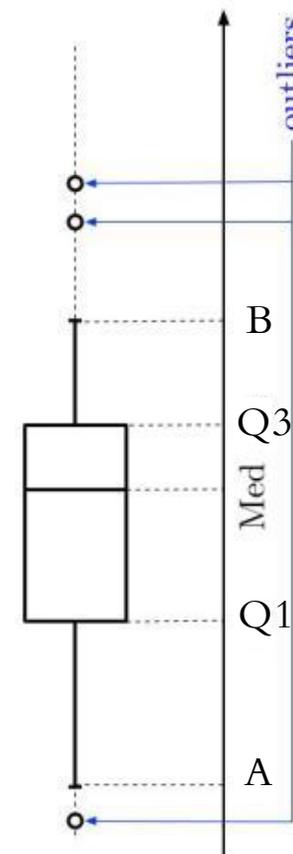
# 3 – Analyse univariées du prix

Méthodes statistiques employées

Répartition des Prix



## Boite à moustache & Ecart interquartile



*Quelques outliers sont apparents (pouvant être des grands crus => valeurs atypiques)*

B – Maximum :  $Q3 + 1,5 \times IQ = 84\text{€}$

Quartile 3 (75%) : 42,03€

**Moyenne des prix : 32,31€**

Ecart type : 27,62€

Quartile 1 (25%) : 14,04€

$IQ = Q3 - Q1$

A – Minimum :  $Q1 - 1,5 \times IQ$



# 4 – Analyses complémentaires

## CA & Top 20



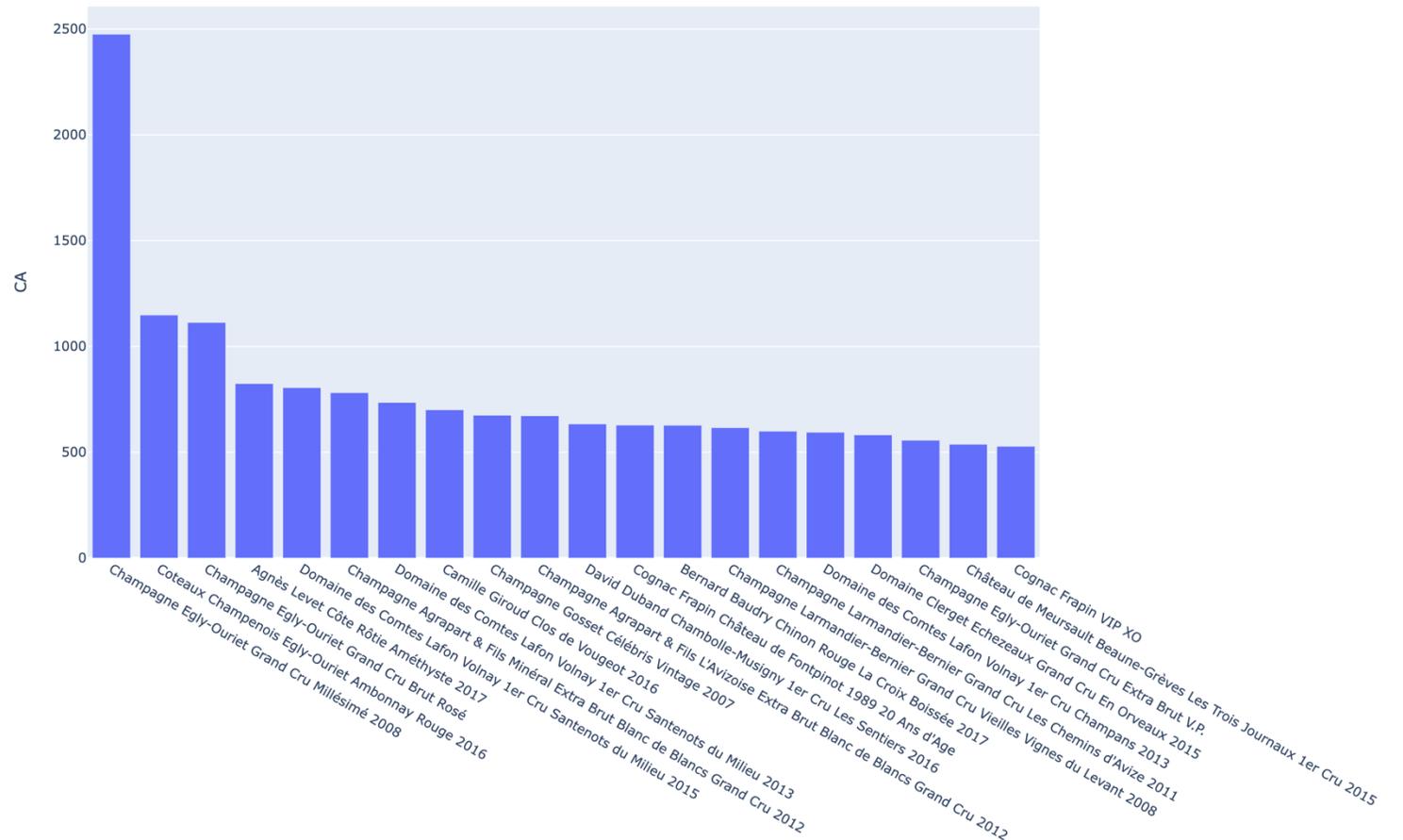
CA du site web

# 143'286 €

```
#Calculez la somme de la colonne "ca_par_article"
ca_par_article_sum = df_erp_liaison_web['ca_par_article'].sum()
#Ce résultat correspond au chiffre d'affaire du site web
print("Le CA du site web est de :", round(ca_par_article_sum))
```

Le CA du site web est de : **143286**

Top 20 Articles par CA





# 4 – Analyses complémentaires

## CA - Commentaires

	sku	total_sales	post_date	product_type	post_title	post_modified	post_type	product_id	onsale_web	price	stock_quantity	stock_status
0	15940	11.0	2018-03-02 10:30:04	Champagne	Champagne Egly-Ouriet Grand Cru Millésimé 2008	2020-03-07 11:18:45	product	4352	1	225.0	0	outofstock
1	14983	6.0	2019-03-28 10:21:36	Champagne	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	2020-04-01 09:30:09	product	5892	1	191.3	98	instock
2	12587	14.0	2018-03-02 10:37:26	Champagne	Champagne Egly-Ouriet Grand Cru Brut Rosé	2020-08-22 11:45:02	product	4353	1	79.5	127	instock
3	15325	20.0	2019-03-27 17:59:49	Vin	Agnès Levat Côte Rôtie Améthyste 2017	2020-05-21 14:00:02	product	5826	1	41.2	34	instock
4	13996	7.0	2019-07-25 09:09:17	Vin	Domaine des Comtes Lafon Volnay 1er Cru Santen...	2020-06-16 09:30:16	product	6212	1	115.0	16	instock
5	13913	9.0	2018-07-18 10:46:30	Champagne	Champagne Agrapart & Fils Minéral Extra Br...	2020-05-11 14:35:02	product	5026	1	86.8	101	instock
6	11602	7.0	2018-07-17 10:52:41	Vin	Domaine des Comtes Lafon Volnay 1er Cru Santen...	2020-06-23 15:35:02	product	5008	1	105.0	12	instock



Point de vigilance sur le  
TOP 1 du CA qui n'est  
plus en stock



## 4 – Analyses complémentaires

CA – 20/80



**432**

articles  
représentent  
**80% du CA**



soit **60%**  
de l'ensemble du  
catalogue



# 4 – Analyses complémentaires

## Quantités & Top 20



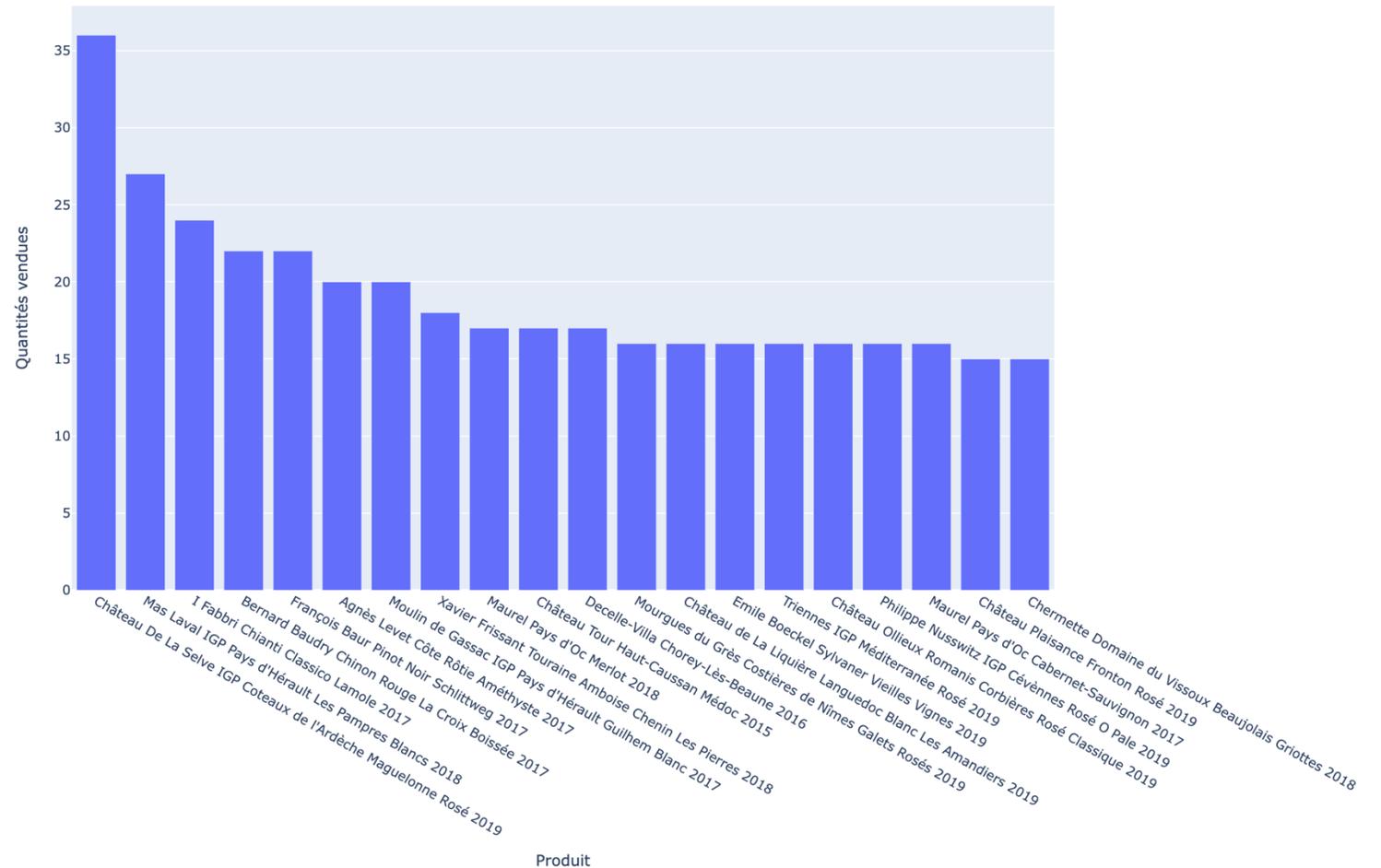
# 5'740

articles vendus  
au total

```
ca_par_article_sum = df_erp_liaison_web['total_sales'].sum()
#Nombre total d'articles vendus
print("Le nombre total d'articles vendus sur le site web est de :", round(ca_par_article_sum))
```

Le nombre total d'articles vendus sur le site web est de : 5740

Top 20 Articles par Quantités vendues





# 4 – Analyses complémentaires

## Quantité - Commentaires

	sku	total_sales	post_date	product_type	post_title	post_modified	post_type	product_id	onsale_web	price	stock_quantity	stock_status
0	16148	36.0	2018-05-03 13:20:05	Vin	Château De La Selve IGP Coteaux de l'Ardèche M...	2020-08-27 09:30:15	product	4867	1	9.9	121	instock
1	15415	27.0	2018-02-15 14:33:42	Vin	Mas Laval IGP Pays d'Hérault Les Pampres Blanc...	2020-07-11 16:45:03	product	4203	1	9.9	74	instock
2	14864	24.0	2018-02-27 13:33:54	Vin	I Fabbri Chianti Classico Lamole 2017	2020-08-22 14:35:02	product	4275	1	14.9	62	instock
3	16525	22.0	2018-04-17 09:28:58	Vin	Bernard Baudry Chinon Rouge La Croix Boissée 2017	2020-07-31 09:31:39	product	4647	1	28.5	45	instock
4	14950	22.0	2018-04-18 11:53:51	Vin	François Baur Pinot Noir Schlittweg 2017	2020-05-06 11:35:01	product	4726	1	12.7	0	outofstock
5	15325	20.0	2019-03-27 17:59:49	Vin	Agnès Levet Côte Rôtie Améthyste 2017	2020-05-21 14:00:02	product	5826	1	41.2	34	instock



Point de vigilance sur le  
TOP 5 en quantités  
vendues qui n'est plus en  
stock



## 4 – Analyses complémentaires

Quantités – 20/80



**431**

articles  
représentent  
**80% des  
quantités  
vendues**



soit **60%**  
de l'ensemble du  
catalogue



# 4 – Analyses complémentaires

## Stocks & Flop 20



Valeur du stock

# 500k€

pour 16710 articles

```
#Calculer la somme de la colonne "Valorisation_stock_euros"
somme_valorisation_stocks = df_erp_liaison_web['valorisation_des_stocks'].sum() # Calculer la somme de la colonne 'valorisation_des_stocks'

# Afficher la somme
print(f"La somme de la valorisation des stocks en euros est de {somme_valorisation_stocks:2f}€")

La somme de la valorisation des stocks en euros est de 493689.60€

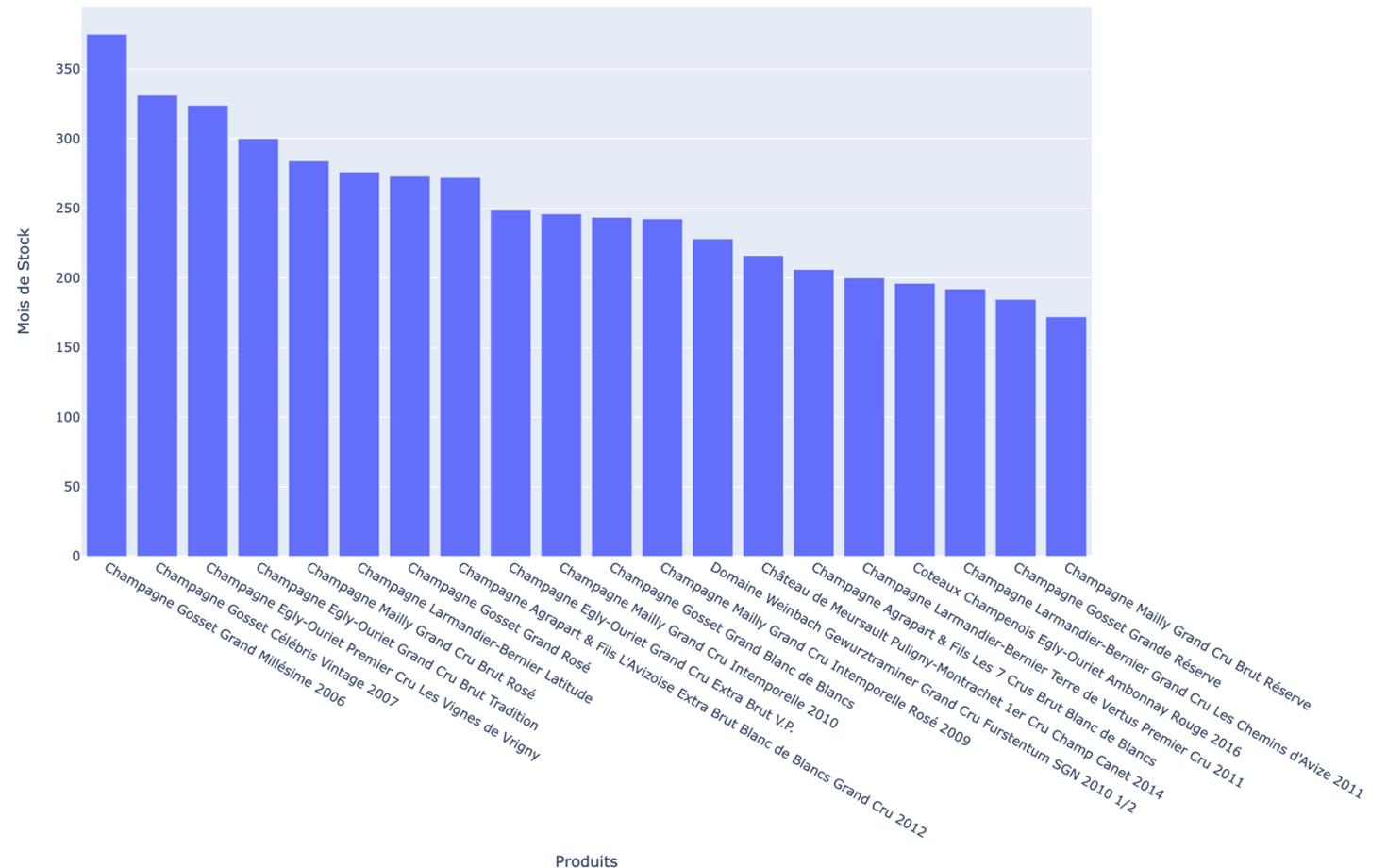
#####
# Valorisation du nombre de produit en stock #
#####

#Calculer la somme de la colonne stock quantity
somme_stock_quantity = df_erp_liaison_web['stock_quantity'].sum()

print("Le niveau de stock est de", somme_stock_quantity, "produits.")

Le niveau de stock est de 16710 produits.
```

Flop 20 des produits avec le plus de mois de stock





## 4 – Analyses complémentaires

Taux de marge



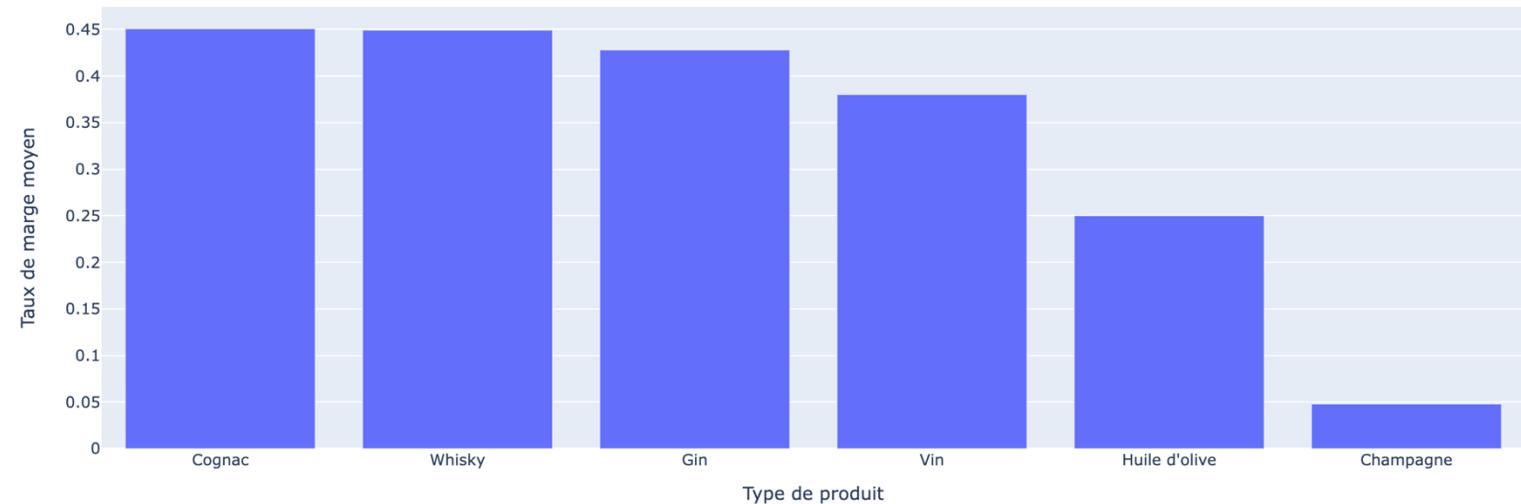
### Marges

Max : 48%

Min : 23%

*Certains produits vendus à perte ont fait plonger la marge sur les Champagnes*

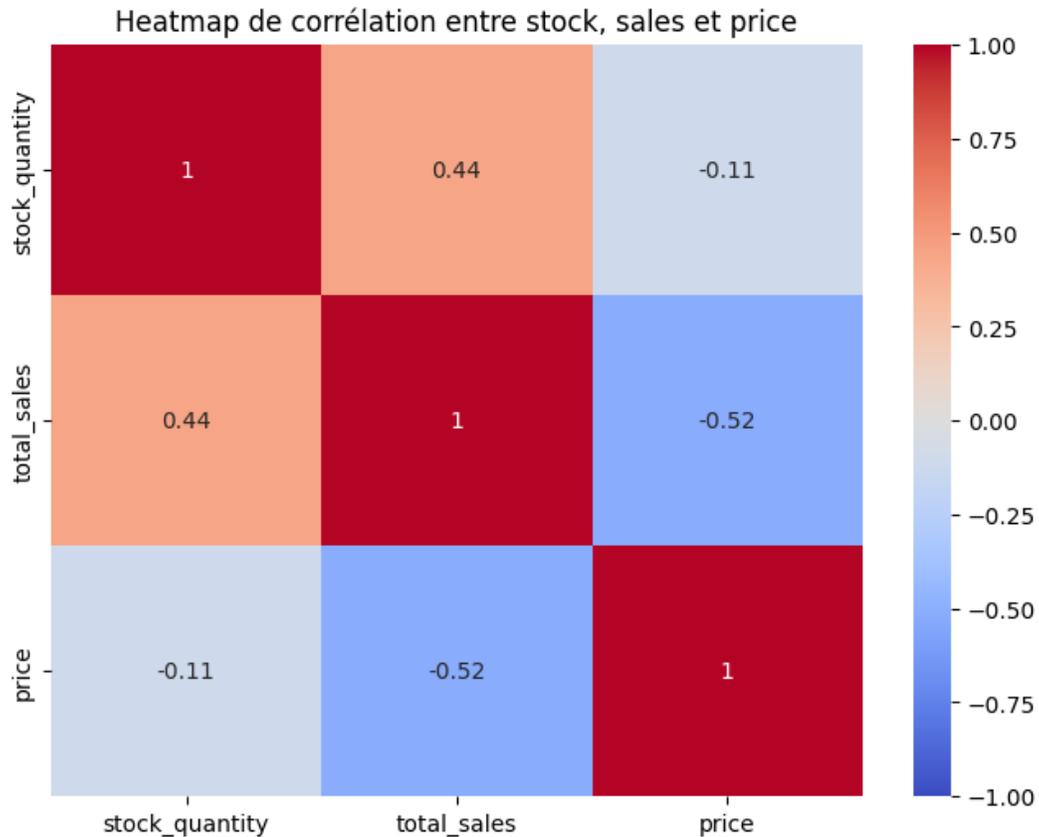
Taux de marge moyen par type de produit





# 4 – Analyses complémentaires

## Corrélations



Les valeurs annotées indiquent **le degré de corrélation** (les valeurs proches de 1 indiquent une corrélation positive forte, proche de -1 une corrélation négative forte, et proche de 0 une faible corrélation) :

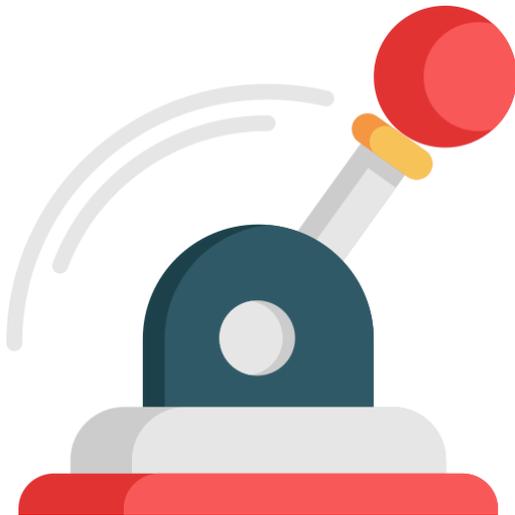
- Le stock est donc significativement corrélé aux ventes
- Les ventes ne sont pas corrélées au prix
- Les prix sont très faiblement corrélés au stock



## 5 – Actions pour la suite

### Produits

Dans les actions à entreprendre, on pourrait :



- **Pousser l'analyse sur la marge des Champagnes** qui est anormalement faible (vente à perte de certains produits).
- **S'assurer d'une bonne corrélation entre les stocks et les ventes de produits.**  
Envisager éventuellement des offres promotionnelles sur les produits du Flop 20 pour réduire l'immobilisation du stock.  
Eviter les ruptures de stock, surtout sur les produits les plus vendus (Top 1 en CA et Top 5 en quantité)



## 5 – Actions pour la suite

Données ERP



- **La colonne « stock\_status »** est corrélée avec « stock\_quantity ». **Peut se poser la question de conserver cette information** (d'autant que si cette mention est entrée manuellement, cela peut générer des erreurs).
- **Certains prix de vente avaient une valeur négative, ce qui n'est pas possible.** Il pourrait être pertinent de bloquer les valeurs négatives dans les prix de vente de l'ERP pour éviter les erreurs de signe.



## 5 – Actions pour la suite

Données Web & RGPD



- Les **données web comportaient beaucoup de caractéristiques** (colonnes) dont nombreuse n'étaient pas utiles à l'analyse visée.  
Dans un objectif d'efficacité et de respect du RGPD (article 5), toutes les données inutiles au traitement ont été écartées.
- **Des erreurs étaient présentes dans le 'sku'**, valeurs vides ou erronées (nan '13127-1' 'bon-cadeau-25-euros').  
Il serait pertinent d'avoir une génération automatisée et incrémentale de cet identifiant sur la base d'un masque prédéfini avec aucune possibilité de modification manuelle, ce qui éviterait les erreurs rencontrées.



## 5 – Point sur les compétences apprises

--



- **Découverte & utilisation** de Jupyter Notebook et des librairies python pour la data analyse (Pandas, Plotly, Matplotlib, plotly express, numpy, seaborn)
- **Manipulation** des data frame (Pandas)
- **Utilisation** de la Data Visualisation (Matplotlib.pyplot, plotly express, seaborn)
- **Méthodologie** d'analyse (notebook guidé)



**END**

Merci pour votre écoute 😊