

# OPTIMISER LA GESTION DES DONNÉES D'UNE BOUTIQUE AVEC PYTHON



Mame Diarra **DABO**

31/03/2022

# PROBLÉMATIQUE

Data Analyst



L'objectif est « rapprochement entre les deux bases »

## Mission: Demandes de Laurent

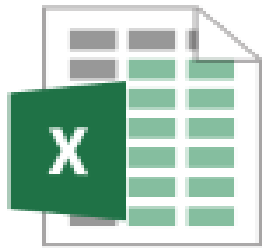
- Rapprocher deux exports ;
- Le chiffre d'affaires par produit, ainsi que le total du chiffre d'affaire réalisé en ligne ;
- La vérification des erreurs de saisie dans certains prix des produits ;
- Effectuer une analyse sur cette variable afin de détecter des valeurs aberrantes.



# PLAN

- Importer le jeu de données
- Nettoyages des données
- Analyse exploratoire
- Répondre aux demandes

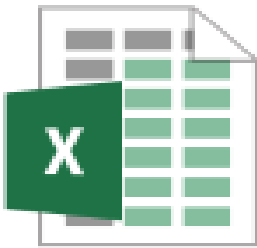
# LE JEU DE DONNÉES



liaison.xlsx



web.xlsx



erp.xlsx

*3 fichiers.xlsx:*

- *Fichier ERP : (825, 5)*
- *Fichier web : (1513, 28)*
- *Fichier liaison : (825, 2)*

# NETTOYAGES DES DONNÉES

Afficher chaque jeu de données

- Nombre de lignes et colonnes :
  - => les types de données
  - => Vérifier les doublons
  - => Vérifier les valeurs manquantes
- Gestion des valeurs
  - => Convertir au bon format de données
  - => Travailler avec les valeurs existantes

# ANALYSE EXPLORATOIRE

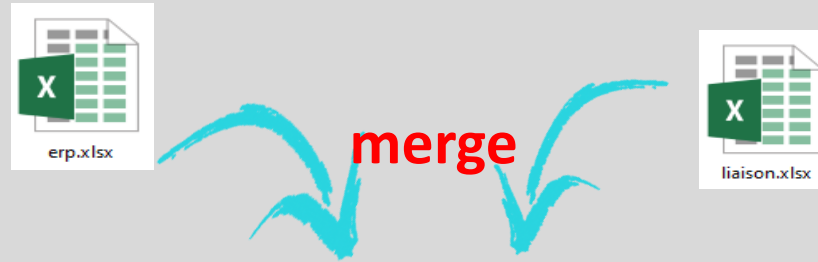
	count	mean	std	min	25%	50%	75%	max
onsale_web	825.0	0.869091	0.337506	0.0	1.0	1.0	1.0	1.0
price	825.0	32.415636	26.795849	5.2	14.6	24.4	42.0	225.0
stock_quantity	825.0	26.496970	45.892422	0.0	1.0	11.0	34.0	578.0

	count	mean	std	min	25%	50%	75%	max
virtual	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
downloadable	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
rating_count	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
average_rating	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
total_sales	714.0	4.012605	8.518183	0.0	0.0	1.0	4.0	96.0
tax_class	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
post_author	714.0	1.998599	0.037424	1.0	2.0	2.0	2.0	2.0
post_content	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
post_password	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
post_content_filtered	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
post_parent	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
menu_order	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
comment_count	714.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0

Utilisation de la fonction `describe()` pour obtenir une description statistique de chaque variable.

# Demandes de Laurent

## Rapprochement des deux exports



Un fichier avec : 825 références de produits

7 colonnes

- **Résultat**



Un fichier avec : 825 références de produits en ligne

35 colonnes

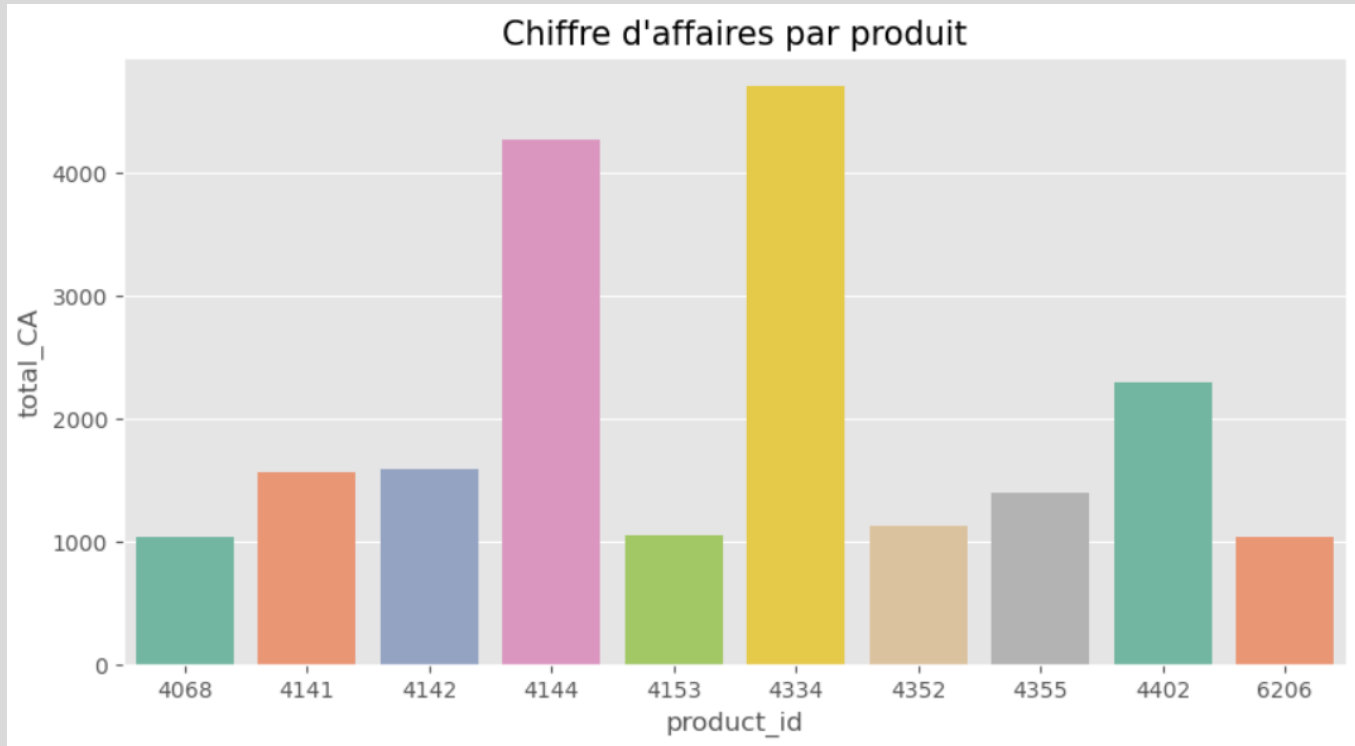


**Fichier final :** 714 références de produits en ligne

34 colonnes

# Demandes de Laurent

## Chiffre d'affaire par produit



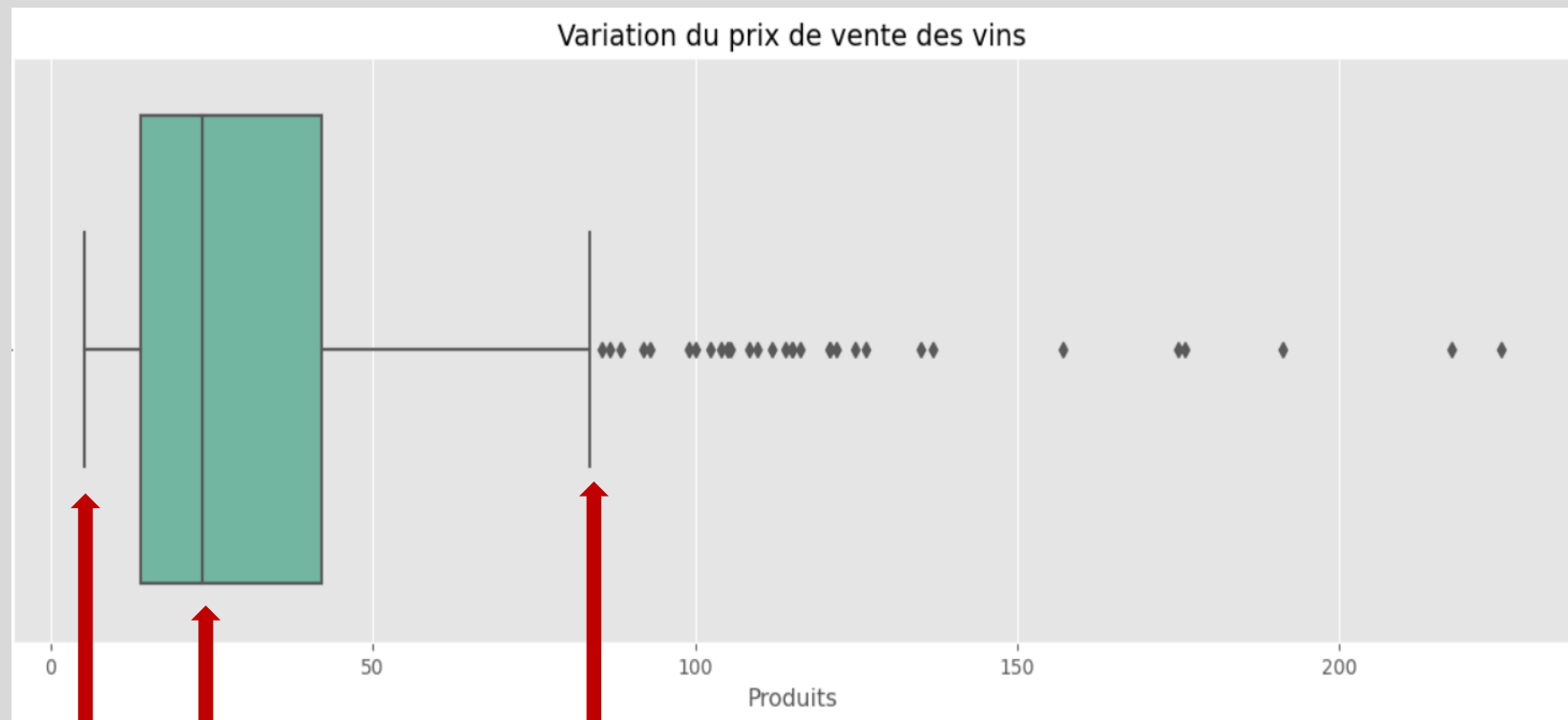
```
Tot_chiff_daff_lign = chiff_daff_prod["total_CA"].sum()  
print(f" Le total du chiffre d'affaires réalisé en ligne est de : {Tot_chiff_daff_lign} euros")
```

Le total du chiffre d'affaires réalisé en ligne est de : 70568.6 euros



# Demandes de Laurent

## Analyse de la variable « prix » sous forme de boxplot



Prix min ~ 5 euros

Prix max ~ 82 euros

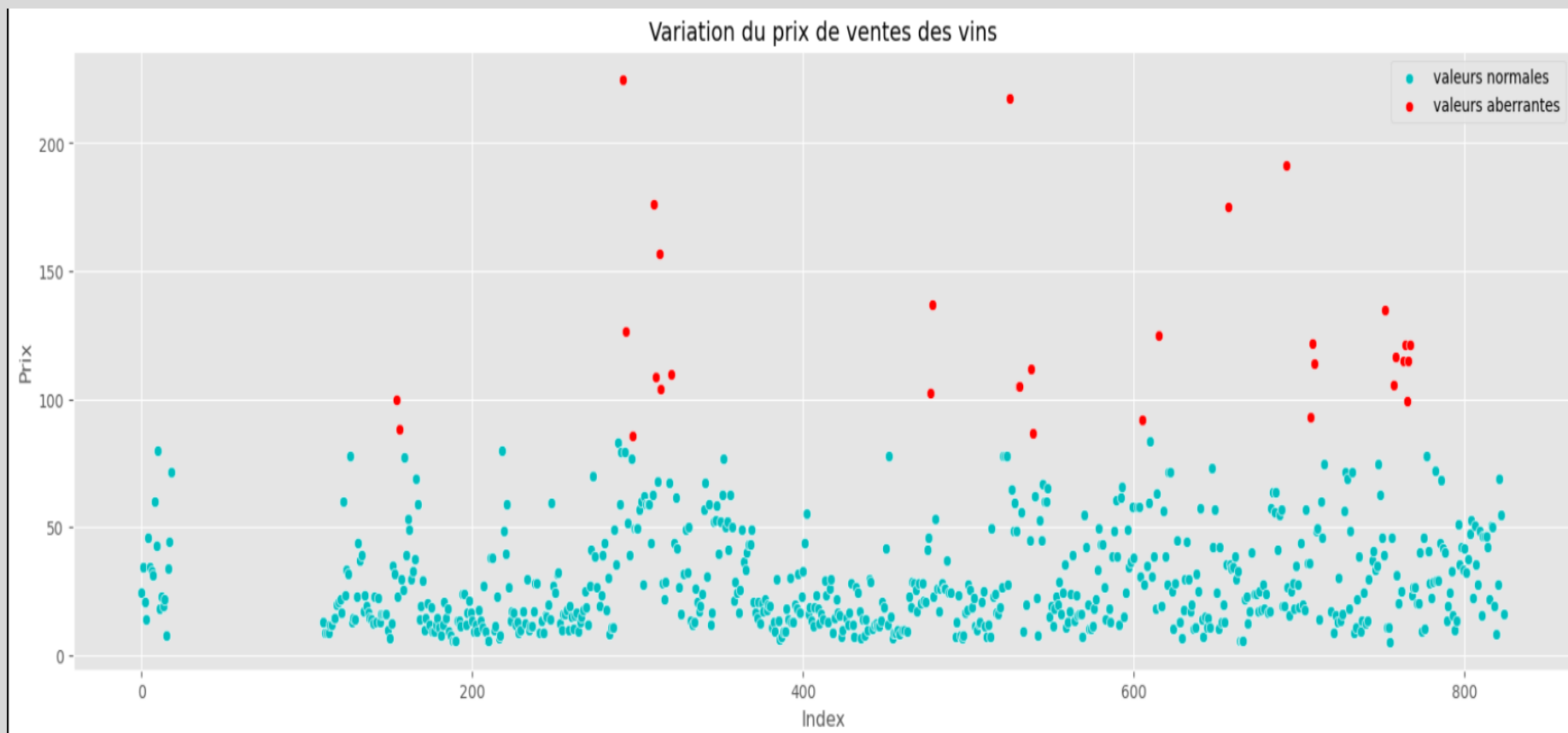
Prix médian ~ 22 euros

(32, 1)

	price
154	100.0
156	88.4
291	225.0
293	126.5
297	85.6

# Demandes de Laurent

Analyse de la variable « prix » sous forme de nuages de points



- En rouge les outliers : ●
- En vert les prix normales : ●

# AUTRES

## Piste à améliorer

- Gestion des Valeurs manquantes
- Traiter les valeurs aberrantes
- Faire plus de visualisation

# BILAN



- Données



- Explore et Analyse



- Réponses

