



# Rapprochement et analyse de nos données

- **Réaliser un état des lieux des données dont nous disposons**
  - **Nettoyer les données**
  - **Consolider les données**
  - **Analyser les données**
  - **Recommandations**

## 3 datasets

➤ ERP

➤ Site Web

➤ Table de liaison

# Analyses Exploratoires des Données



## ERP

3 types d'erreurs

| product_id | onsale_web | price | stock_quantity | stock_status | purchase_price | stock_status_2 |
|------------|------------|-------|----------------|--------------|----------------|----------------|
| 4039       | 1          | 46.0  | 3              | outofstock   | 23.77          | instock        |
| 4885       | 1          | 18.7  | 0              | instock      | 9.66           | outofstock     |

|      |   |      |     |            |       |         |
|------|---|------|-----|------------|-------|---------|
| 4973 | 0 | 10.0 | -10 | outofstock | 4.96  | instock |
| 5700 | 1 | 44.5 | -1  | outofstock | 22.30 | instock |

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

# Analyses Exploratoires des Données



## WEB

5 types d'erreurs

- 85 lignes sans aucune donnée
- 714 doublons sur la clé primaire

| sku | total_sales | product_type | post_title                                      | post_name                                  | post_date_gmt       | post_modified_gmt   | post_author | post_type | post_excerpt                                      | guid                                              |
|-----|-------------|--------------|-------------------------------------------------|--------------------------------------------|---------------------|---------------------|-------------|-----------|---------------------------------------------------|---------------------------------------------------|
| NaN | -56.0       | Vin          | Pierre Jean Villa Condrieu Jardin Suspendu 2018 | pierre-jean-villa-condrieu-suspendu-2018   | 2018-08-08 09:23:43 | 2019-11-02 12:24:01 | 2.0         | product   | <span id="u1194-83">Le nez séduit par ses parf... | https://www.bottle-neck.fr/?post_type=product&... |
| NaN | -17.0       | Vin          | Pierre Jean Villa Côte Rôtie Fongeant 2017      | pierre-jean-villa-cote-rotie-fongeant-2017 | 2018-07-31 10:07:23 | 2019-11-02 12:24:15 | 2.0         | product   | Fongeant 2017 explose sur un fruit brillant, p... | https://www.bottle-neck.fr/?post_type=product&... |

Valeurs sku non conformes à la longueur de 6 caractères :

```
272          13127-1
842    bon-cadeau-25-euros
1084          SKU_null1
1087          SKU_null2
1117          13127-1
1387    bon-cadeau-25-euros
```

| sku                 | total_sales | product_type | post_title        | post_name              | post_date_gmt       | post_modified_gmt   | post_author | post_type | post_excerpt                                      | guid                                              |
|---------------------|-------------|--------------|-------------------|------------------------|---------------------|---------------------|-------------|-----------|---------------------------------------------------|---------------------------------------------------|
| bon-cadeau-25-euros | 7.0         | NaN          | Bon cadeau de 25€ | bon-cadeau-de-25-euros | 2018-06-01 11:53:46 | 2018-06-01 12:13:57 | 1.0         | product   | <span style="color: #a85253;"><strong>Parlons ... | https://www.bottle-neck.fr/?post_type=product&... |

## LIAISON

2 types d'erreurs

- 91 valeurs manquantes
- 18 valeurs supplémentaires dans web

|                                                 |     |        |     |
|-------------------------------------------------|-----|--------|-----|
| Nombre de valeurs uniques pour chaque colonne : | {}  | id_web | 734 |
| product_id                                      | 825 |        |     |

Valeurs id\_web non conformes à la longueur de 5 caractères :

|     |                     |
|-----|---------------------|
| 443 | bon-cadeau-25-euros |
| 822 | 13127-1             |
| 823 | 14680-1             |

## Jonction des fichiers ERP et LIAISON

➤ Les 825 clés product\_id correspondent entre les tables

| product_id | price | purchase_price | onsale_web | stock_quantity | stock_status | id_web |
|------------|-------|----------------|------------|----------------|--------------|--------|
| 4233       | NaN   | 10.33          | 0          | 0              | outofstock   | NaN    |
| 5017       | NaN   | 4.34           | 0          | 0              | outofstock   | NaN    |
| 6594       | NaN   | 4.61           | 0          | 19             | instock      | NaN    |

## Jonction des fichiers avec WEB

- Il y a 716 clés id\_web dans WEB
- Il y a 734 clés id\_web dans LIAISON

### Données non exploitables :

- Ces 18 « nouvelles clés »
- Les 2 clés fictives « SKU\_NULL »
- Les 109 produits non référencés dans la table WEB



## Vérifications après jonction

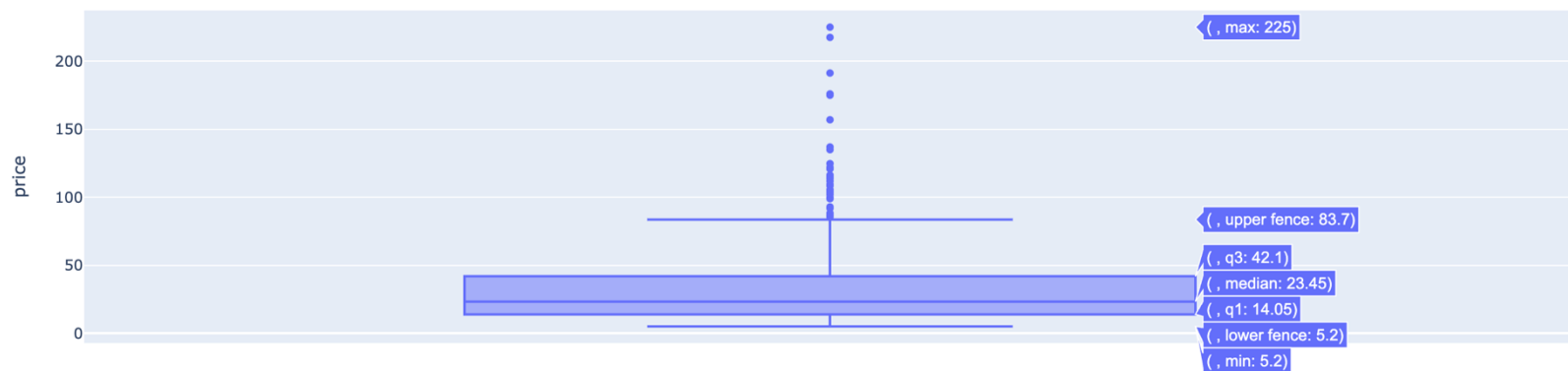
2 anomalies

| product_id | price | purchase_price | onsale_web | stock_quantity | stock_status | id_web | total_sales | product_type | post_title                                              | post_name                                                 | post_date_gmt          | post_modified_gmt      | post_author | post_type |
|------------|-------|----------------|------------|----------------|--------------|--------|-------------|--------------|---------------------------------------------------------|-----------------------------------------------------------|------------------------|------------------------|-------------|-----------|
| 4200       | 5.8   | 3.12           | 0          | 33             | instock      | 16295  | 14.0        | Vin          | Moulin de Gassac<br>IGP Pays d'Hérault<br>Guilhem Ro... | moulin-de-gassac-<br>igp-pays-dherault-<br>guilhem-ros... | 2018-02-15<br>13:05:06 | 2020-08-27<br>16:55:03 | 2.0         | product   |

| product_id | price | purchase_price | onsale_web | stock_quantity | stock_status | id_web                          | total_sales | product_type | post_title              | post_name                      | post_date_gmt          | post_modified_gmt   | post_author | post_type |
|------------|-------|----------------|------------|----------------|--------------|---------------------------------|-------------|--------------|-------------------------|--------------------------------|------------------------|---------------------|-------------|-----------|
| 4954       | 25.0  | 13.3           | 1          | 23             | instock      | bon-<br>cadeau-<br>25-<br>euros | 7.0         | NaN          | Bon<br>cadeau<br>de 25€ | bon-<br>cadeau-de-<br>25-euros | 2018-06-01<br>11:53:46 | 2018-06-01 12:13:57 | 1.0         | product   |

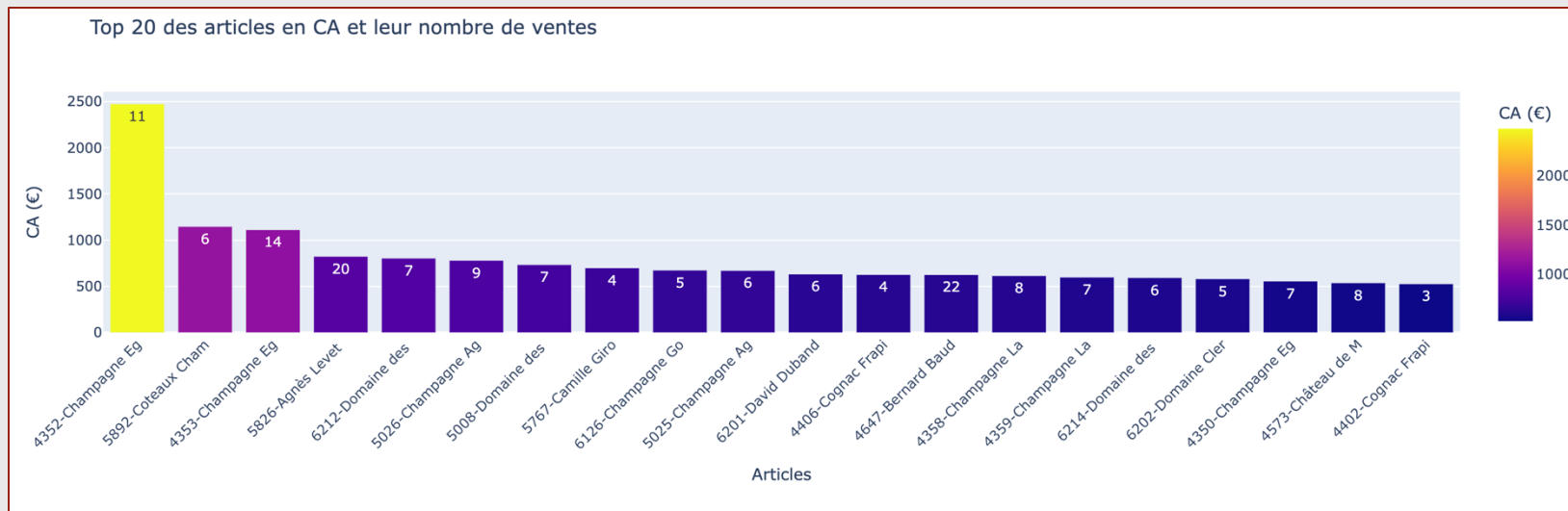
# Analyses univariées du prix

Répartition des prix



# Analyses complémentaires

## Chiffre d'affaires



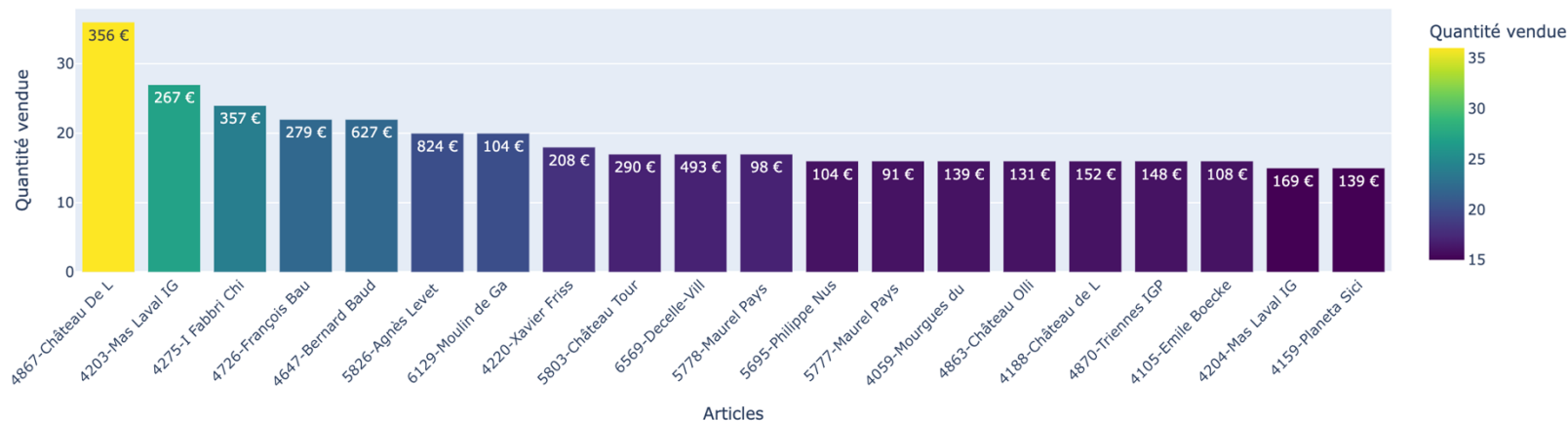
Le champagne représente près de la moitié du chiffre d'affaires du TOP 20  
mais moins de 10% du CA total

# Analyses complémentaires

## Quantités vendues



Top 20 des articles en quantité et leur CA



Le vin représente 94,87 % du nombre de ventes

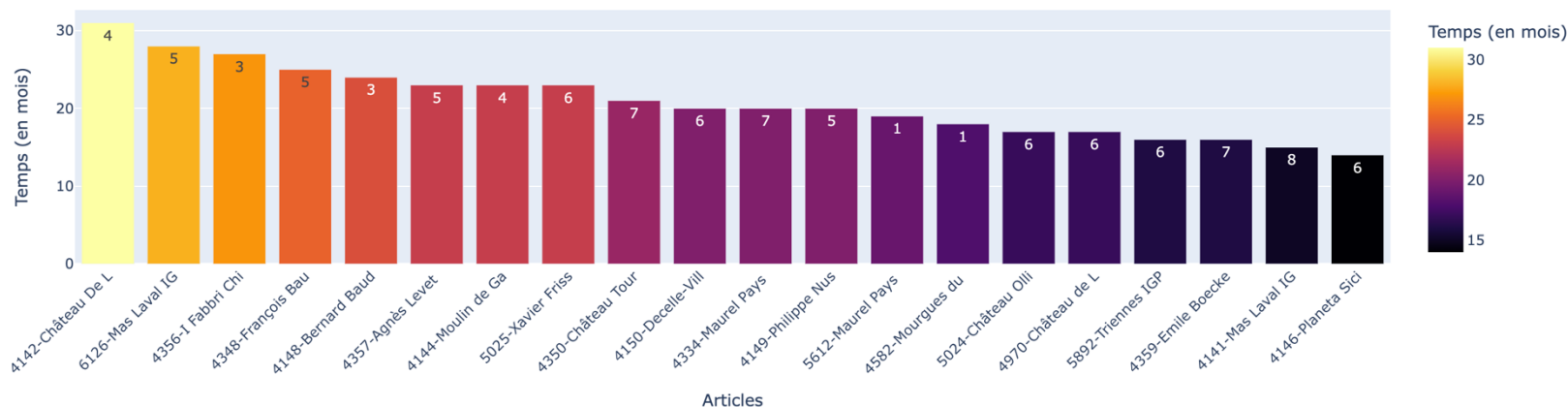
Le champagne en représente 2,94 %

# Analyses complémentaires

## Stocks



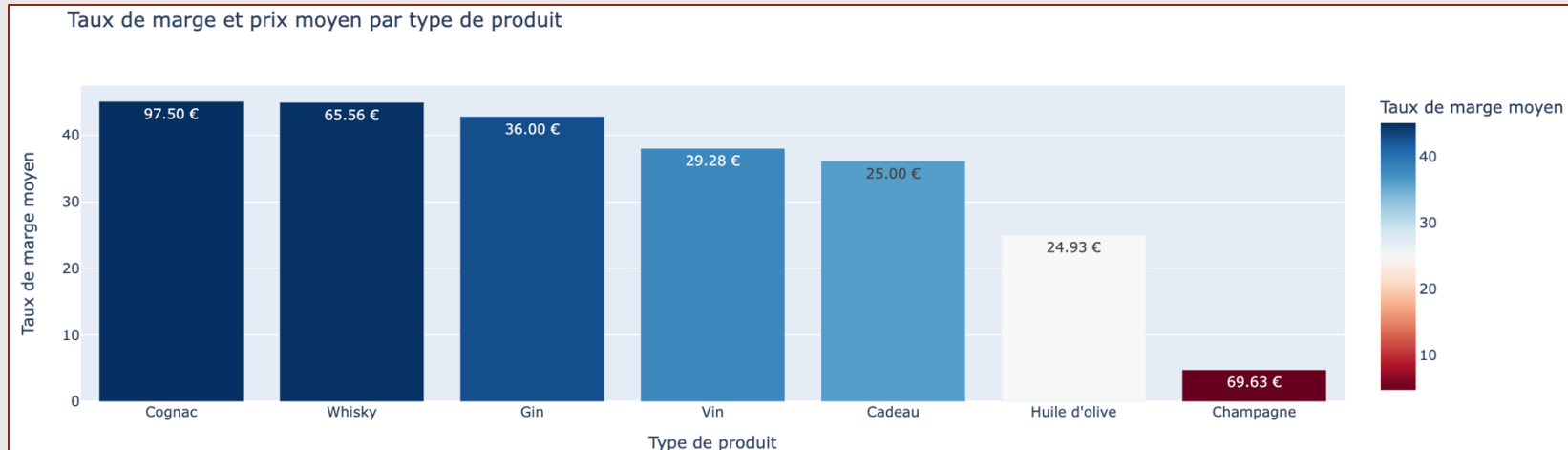
Flop20 des pires rotations de stock et leur nombre de vente



Le champagne représente 38,91 % de la valorisation du stock  
pour seulement 2,94 % du nombre de ventes

# Analyses complémentaires

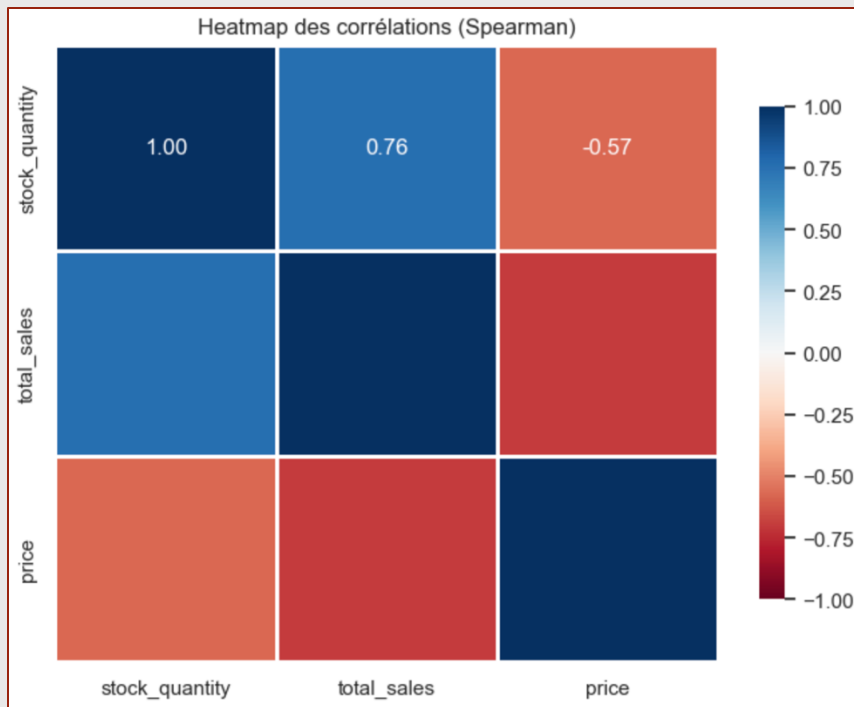
## Taux de marge



Le champagne représente 9 % du chiffre d'affaires mais sa marge n'est que de 4,79 %  
Il semble y avoir une erreur de saisie pour un produit ayant un taux de marge de -635.10 %

# Analyses complémentaires

## Corrélations



Il y a une très forte corrélation entre le nombre de ventes et les quantités de stock

Il y a aussi une corrélation négative assez forte entre le nombre de ventes et le prix des produits

Cela signifie que les ventes baissent lorsque les prix sont élevés

## Pour les datasets

### ERP

- ✓ Créer une colonne total\_sales

Afin que l'on puisse intégrer les ventes hors site web

### WEB

- ✓ Supprimer les colonnes inutiles
- ✓ Mettre à jour les nouveaux produits vendus sur le web
- ✓ Intégrer les posts « attachement » ailleurs

Afin d'éviter de nombreux doublons

### LIAISON

- ✓ Mettre à jour régulièrement les clés

Afin de garantir et d'optimiser la liaison entre les tables



## Pour les données en général

- ✓ Renseigner chaque **élément** et ce dès que possible
- ✓ Faire attention aux **erreurs de saisie**
- ✓ Faire attention aux **données improbables**
- ✓ Faire remonter toute **anomalie**

## Attention aux méthodes d'analyse parfois arbitraires

- Z-score
- Ecart interquartile
- Loi de Pareto 80/20
- Corrélation de Pearson



**Merci pour votre attention**