



Analyse CA Web

Soutenance OC P5 - 23/05/2022- 10h00

Vincent ARESE

Analyse CA Web Bottleneck

• Sommaire

1. Nettoyage du jeu de données
2. Consolidation des deux Exports (web et Erp)
3. Chiffre d'Affaires boutique Web
4. Recherche éventuelles valeurs aberrantes

Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?

Nettoyage du jeu de données

- fichier : Erp -> product_id clé candidate

Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?

Recherche valeurs uniques : `df` (825 Lignes, 5 Colonnes)

`product_id` : 825 rows |Nb unique: 825 |Nb doublons: 0 - 0.0 % **Clé candidate** | int64

`onsale_web` : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | int64

`price` : 825 rows |Nb unique: 381 |Nb doublons: 444 - 54.0 % | float64

`stock_quantity` : 825 rows |Nb unique: 128 |Nb doublons: 697 - 84.0 % | int64

`stock_status` : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | object

	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

Nettoyage du jeu de données

- fichier : Liaison - Suppression des NaN (Id_Web)

DataFrame initial:

Recherche valeurs uniques : **df** (825 Lignes, 2 Colonnes)

product_id : 825 rows |Nb unique: 825 |Nb doublons: 0 - 0.0 % **Clé candidate** | int64

id_web : 825 rows |Nb unique: 734 |Nb doublons: 91 - 11.0 % | object

Il y a 91 doublons { (91, 2) - [nan] } dans la colonne id-web du DataFrame

	product_id	id_web
19	4055	NaN
49	4090	NaN
50	4092	NaN
119	4195	NaN
131	4209	NaN

DataFrame nettoyé:

Recherche valeurs uniques : **df** (734 Lignes, 2 Colonnes)

product_id : 734 rows |Nb unique: 734 |Nb doublons: 0 - 0.0 % **Clé candidate** | int64

id_web : 734 rows |Nb unique: 734 |Nb doublons: 0 - 0.0 % **Clé candidate** | object

	product_id	id_web
0	3847	15298
1	3849	15296
2	3850	15300
3	4032	19814
4	4039	19815



Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?

Nettoyage du jeu de données

- fichier : web

DataFrame nettoyé:

Recherche valeurs uniques : **df** (716 Lignes, 5 Colonnes)

sku : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % **Clé candidate** | object

guid_ID : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % **Clé candidate** | int64

total_sales : 716 rows |Nb unique: 41 |Nb doublons: 675 - 94.0 % | float64

post_date : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % **Clé candidate** | datetime64[ns]

post_title : 716 rows |Nb unique: 713 |Nb doublons: 3 - 0.0 % | object

	sku	guid_ID	total_sales	post_date	post_title
0	bon-cadeau-25-euros	4954	10.0	2018-06-01 13:53:46	Bon cadeau de 25€
1	15298	3847	6.0	2018-02-08 12:58:52	Pierre Jean Villa Saint-Joseph Préface 2018
2	15296	3849	0.0	2018-02-08 13:49:41	Pierre Jean Villa Saint-Joseph Rouge Tildé 2017
3	15300	3850	0.0	2018-02-08 14:08:36	Pierre Jean Villa Crozes-Hermitage Accroche Co...
4	19814	4032	3.0	2018-02-09 14:01:05	Pierre Jean Villa IGP Collines Rhodaniennes Ga...
...
762	16135	6930	5.0	2020-04-25 13:22:38	Mouthes Le Bihan Côtes de Duras L'Aimé Chai 2015
767	15891	7023	0.0	2020-05-02 14:53:40	Camin Larredya Jurançon Sec La Virada 2018
768	15887	7025	0.0	2020-05-02 15:00:54	Domaine Jamet Côte Rôtie Fructus Voluptas 2018
797	13127-1	7247	0.0	2020-06-09 15:42:04	Clos du Mont-Olivet Châteauneuf-du-Pape 2007
798	16230	7338	0.0	2020-07-20 11:00:00	Domaine Saint-Nicolas Vin de France Blanc Les ...

716 rows x 5 columns

- Supprimer les doublons
(post_type/Product & post_type/attachement)

- Suppression des
colonnes sans intérêt
pour l'analyse

- Remplacer les codes sku
Manquant par le Guid-Id
(guid_ID : 5070 – 5075

- en attente confirmation code SKU)

Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?

Nettoyage du jeu de données

- fichier : Erp -> product_id clé candidate



Recherche valeurs uniques : `df` (825 Lignes, 5 Colonnes)

`product_id` : 825 rows |Nb unique: 825 |Nb doublons: 0 - 0.0 % **Clé candidate** | int64

`onsale_web` : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | int64

`price` : 825 rows |Nb unique: 381 |Nb doublons: 444 - 54.0 % | float64

`stock_quantity` : 825 rows |Nb unique: 128 |Nb doublons: 697 - 84.0 % | int64

`stock_status` : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | object

	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

Consolidation des Exports web & erp

Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?

```
Recherche valeurs uniques : df ( 825 Lignes, 5 Colonnes)

product_id : 825 rows |Nb unique: 825 |Nb doublons: 0 - 0.0 % Clé candidate | int64
onsale_web : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | int64
price : 825 rows |Nb unique: 381 |Nb doublons: 444 - 54.0 % | float64
stock_quantity : 825 rows |Nb unique: 128 |Nb doublons: 697 - 84.0 % | int64
stock_status : 825 rows |Nb unique: 2 |Nb doublons: 823 - 100.0 % | object
```

	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

erp

DataFrame nettoyé:

```
Recherche valeurs uniques : df ( 716 Lignes, 5 Colonnes)

sku : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % Clé candidate | object
guid_ID : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % Clé candidate | int64
total_sales : 716 rows |Nb unique: 41 |Nb doublons: 675 - 94.0 % | float64
post_date : 716 rows |Nb unique: 716 |Nb doublons: 0 - 0.0 % Clé candidate | datetime64[ns]
post_title : 716 rows |Nb unique: 713 |Nb doublons: 3 - 0.0 % | object
```

	sku	guid_ID	total_sales	post_date	post_title
0	bon-cadeau-25-euros	4954	10.0	2018-06-01 13:53:46	Bon cadeau de 25€
1	15298	3847	6.0	2018-02-08 12:58:52	Pierre Jean Villa Saint-Joseph Préface 2018
2	15296	3849	0.0	2018-02-08 13:49:41	Pierre Jean Villa Saint-Joseph Rouge Tildé 2017
3	15300	3850	0.0	2018-02-08 14:08:36	Pierre Jean Villa Crozes-Hermitage Accroche Co...

web

DataFrame nettoyé:

```
Recherche valeurs uniques : df ( 734 Lignes, 2 Colonnes)

product_id : 734 rows |Nb unique: 734 |Nb doublons: 0 - 0.0 % Clé candidate | int64
id_web : 734 rows |Nb unique: 734 |Nb doublons: 0 - 0.0 % Clé candidate | object
```

	product_id	id_web
0	3847	15298
1	3849	15296
2	3850	15300
3	4032	19814
4	4039	19815

liaison



Global



Segmentation Produit

- Par Tranche de Prix (à partir de price)

- `Prix > 100€`
- `50€ > Prix < 100€`
- `30€ > Prix <= 50€`
- `10€ > Prix <= 30€`
- `Prix <= 10€`

- Par type de Produits (à partir de post_title)

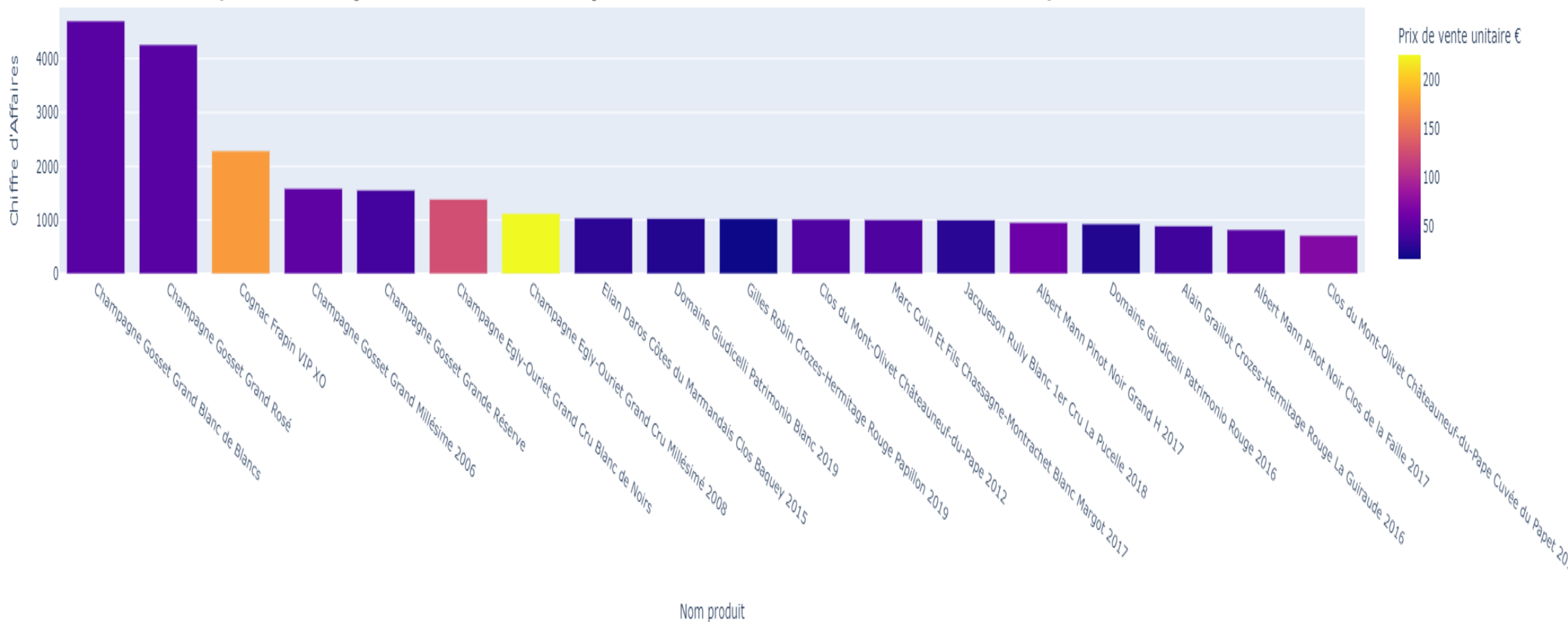
- Champagne
- Cognac
- Whisky
- Autres (vins)



Détail par Sku dynamique :
Cf l'export html ou Excel & Notebook

CA boutique Web

Chiffre d'affaires BottleNeck 70 568.60 €
(sku représentant plus de 1% des ventes)



Consolidation
erp & web

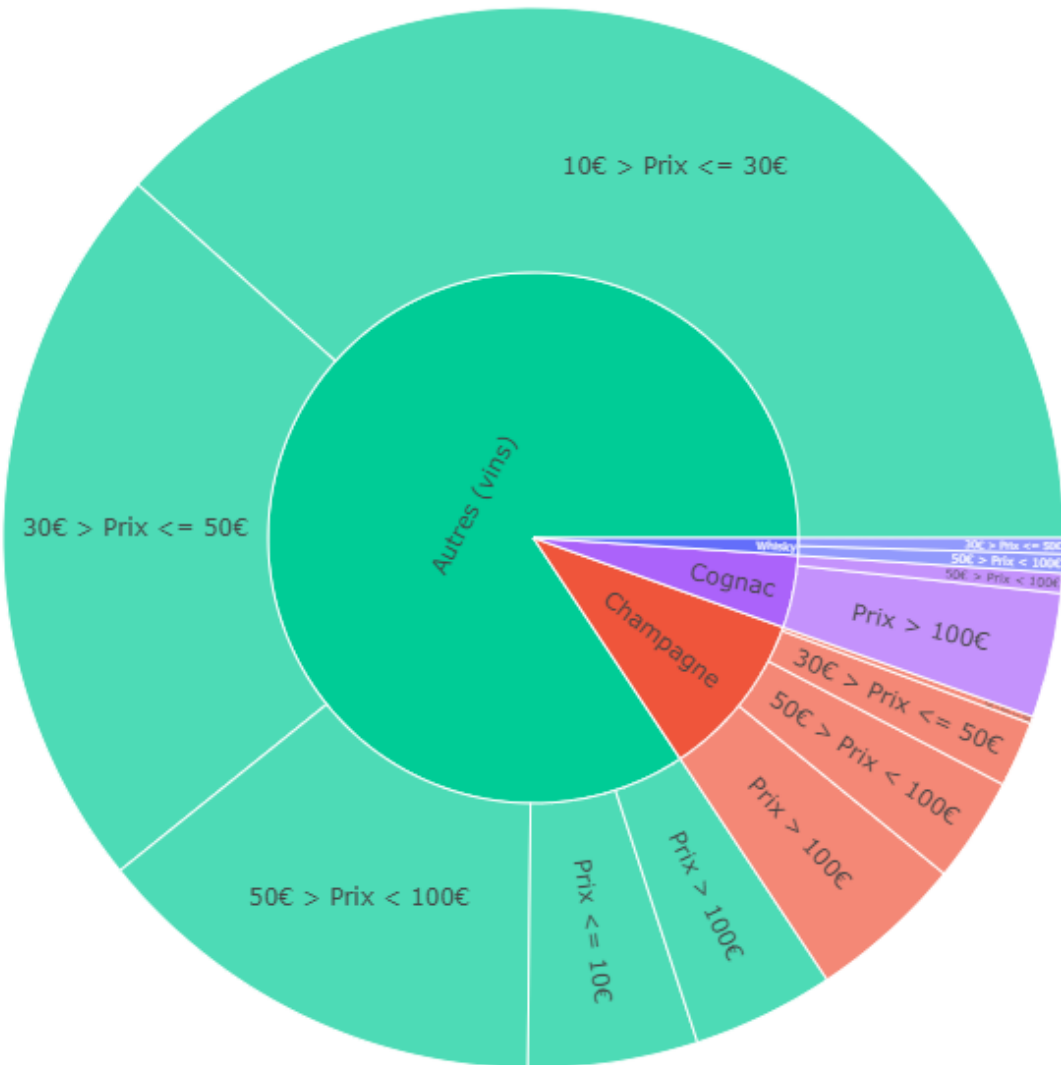
Segmentation
Produit

CA web

Outliers ?

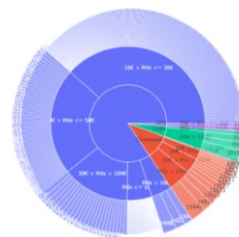


CA boutique Web



Chiffre d'affaires

BottleNeck web : 70 568.60 €
(ventillation par Catégorie produits)



Détail par Sku dynamique :
Cf l'export html ou Excel & Notebook

Recherche Outliers

- **Méthodes utilisées**

- Graphique
- Calcul Z-score
- Méthode Interquartiles

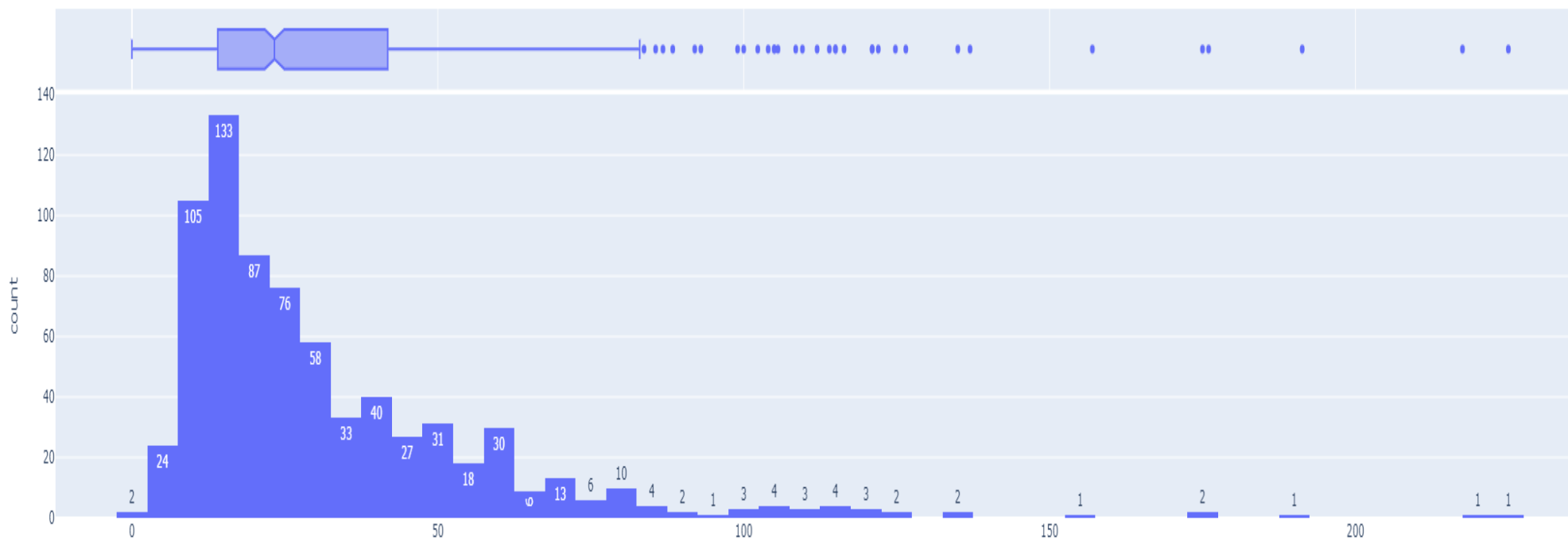
- **Segmentation produits nécessaire** pour une étude plus fine car de nombreux faux-positifs lorsque des produits de nature vraiment différentes sont mélangés (vins , champagne, cognac, whisky)

- Vu les profils de prix et les gammes de produits, les tests statiques (Zscore – interquartiles) permettent de mettre en évidence des outliers possible, mais une levée de doute par un expert métier est nécessaire pour valider si il s’agit effectivement d’outliers.



Recherche Outliers

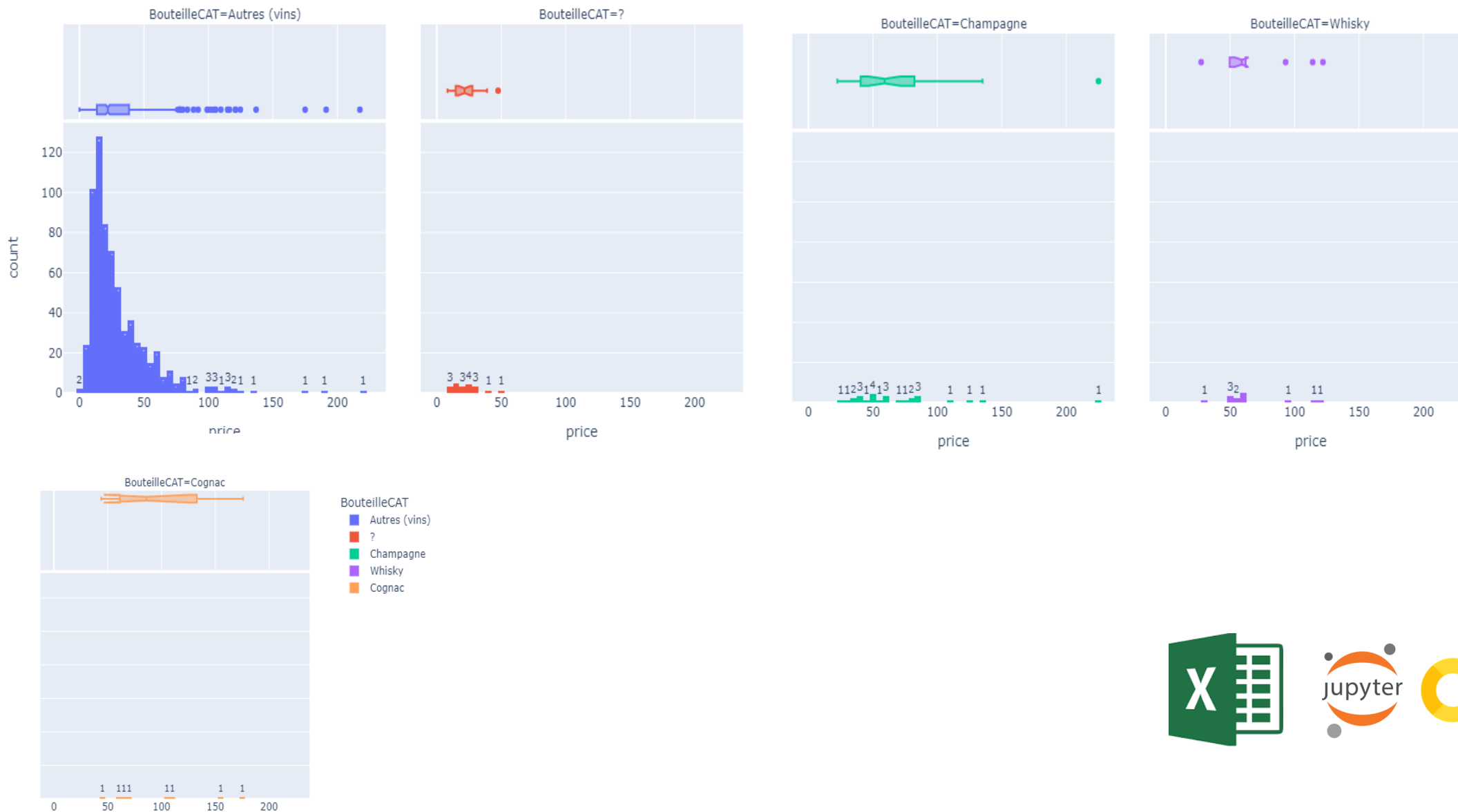
- Distribution des prix unitaires des bouteilles(€)



Répartition des prix unitaires et visualisation des Outliers potentiels & Outliers

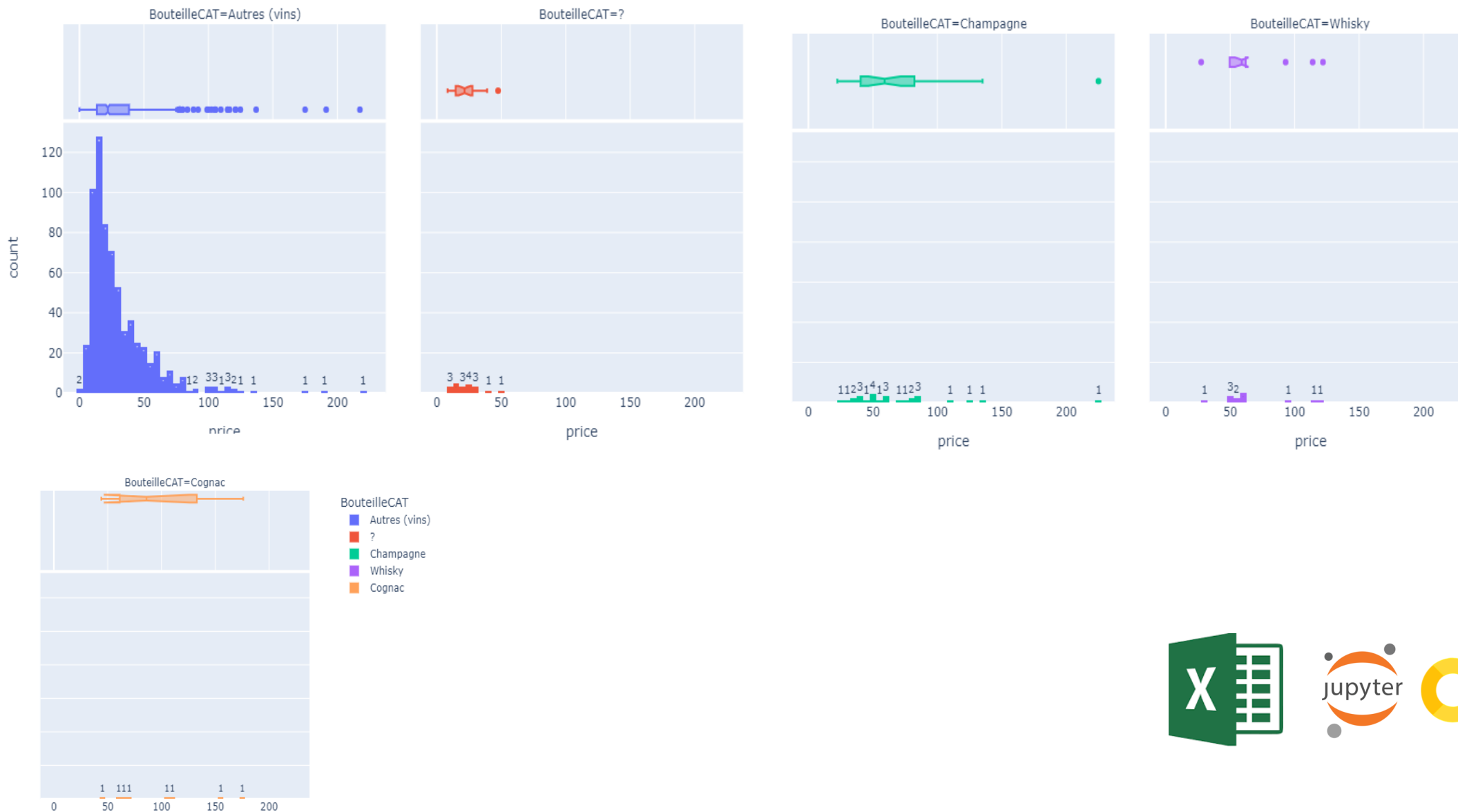
Recherche Outliers

Distribution des prix unitaires des bouteilles € par cateregorie de boissons



Recherche Outliers

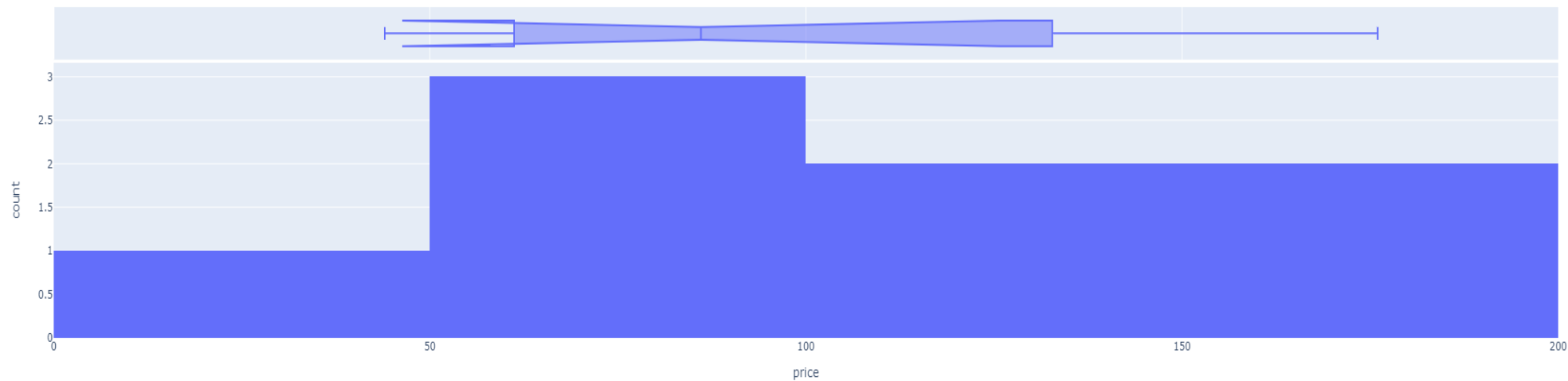
Distribution des prix unitaires des bouteilles € par cateregorie de boissons



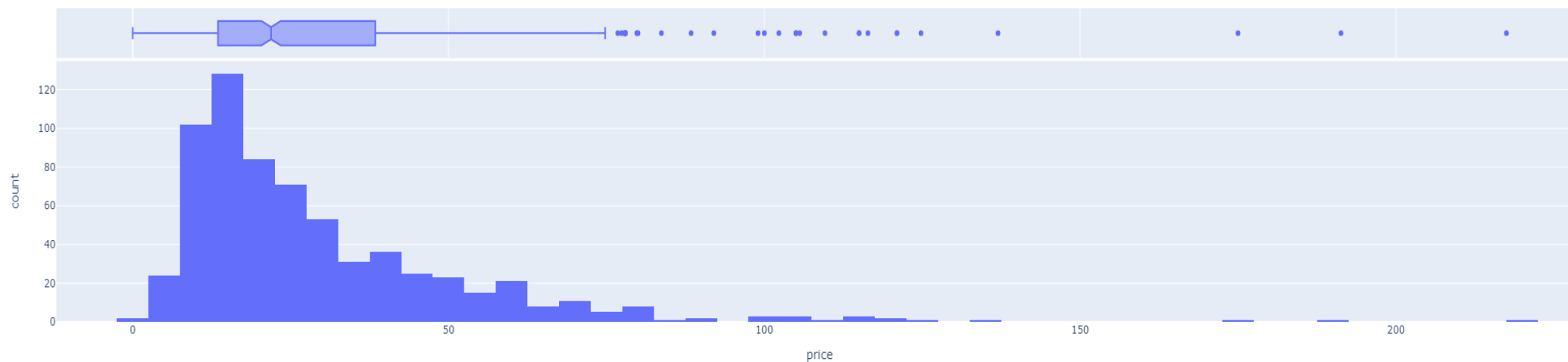
Recherche Outliers : Champagne & vins



Distribution des prix unitaires des bouteilles de Cognac en €

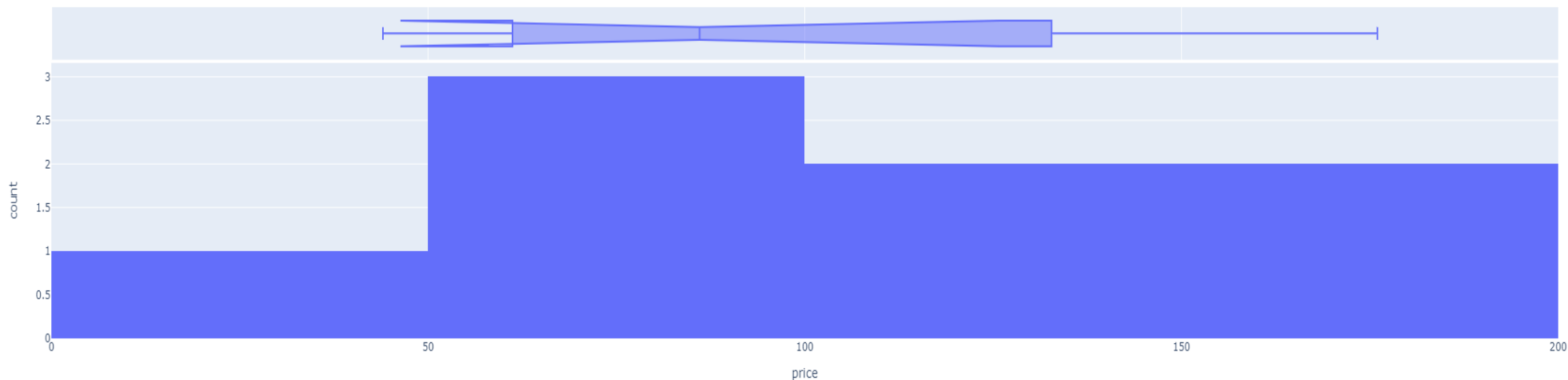


Distribution des prix unitaires des bouteilles de vins (autre) en €

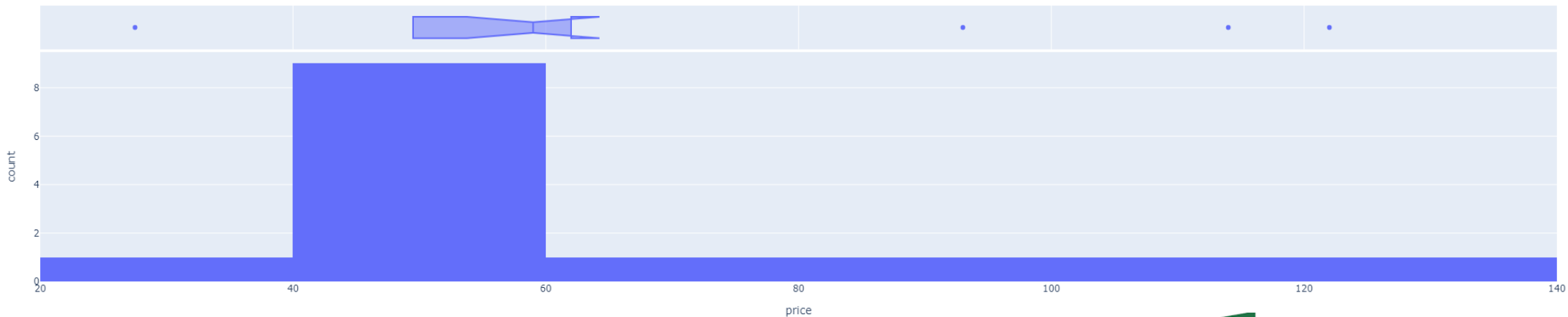


Recherche Outliers : Coanac & Whisky

Distribution des prix unitaires des bouteilles de Cognac en €



Distribution des prix unitaires des bouteilles de Whisky en €



Consolidation
erp & web

Segmentation
Produit

CA web

Outliers ?



Questions

- **Merci pour votre attention**

