

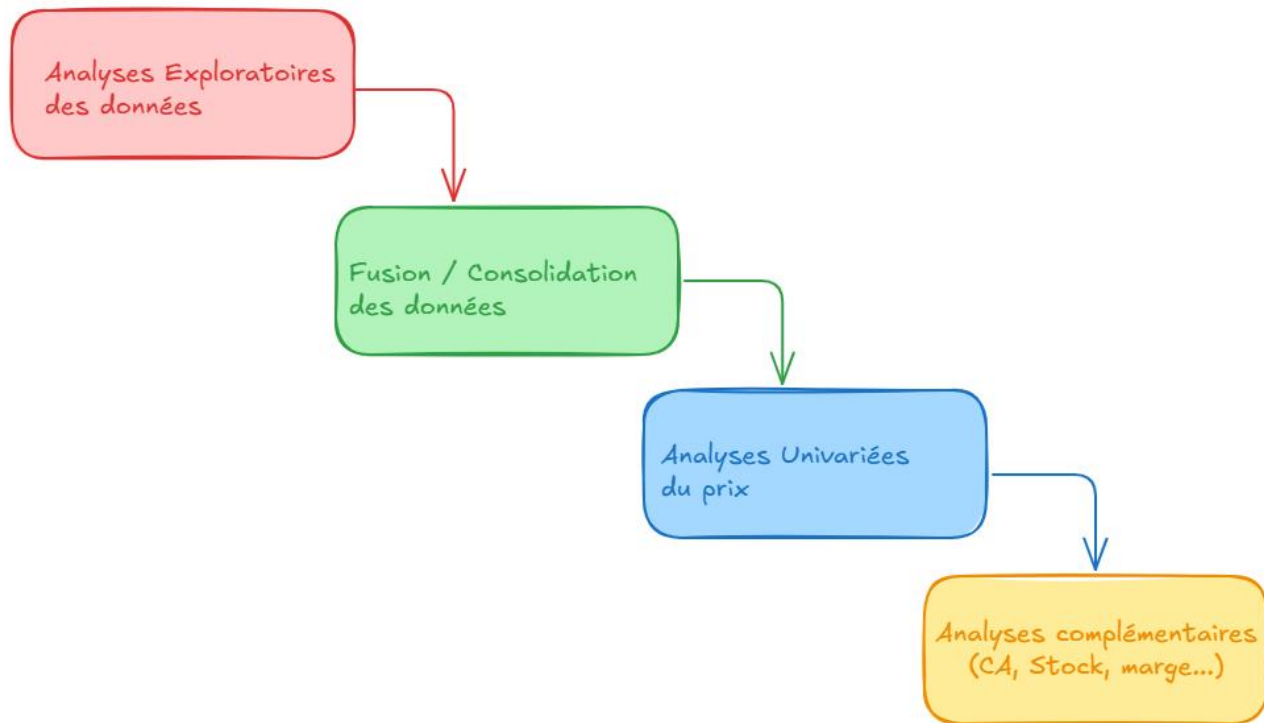
Bottleneck

Poirier Pierre-Marie

Data-analyst

31/10/2024

Méthodologie et Timeline



Analyses Exploratoires des Données

DF_ERP

- *Données issues de l'erp avec toutes les informations centralisées*

DF_Web

- *Données issues de la boutique e-commerce (wordpress)*

DF_Liaison

- *Table de liaison entre erp (product_id) et web(sku)*

Analyses Exploratoires des Données : ERP

DF_ERP

	Colonne 1	...	Colonne 6
Ligne 1			
.			
.			
.			
Ligne 825			

- Table référençant l'état des produits vendus (id, stock, prix d'achat, prix de vente)
Correction des erreurs:
- Fiabilisation de l'infos **instock** et **outstock** via création d'une nouvelle colonne stock_status + comparaison (avec fonction apply si x =0 alors outofstock sinon instock)
- correction des prix négatif qu'on transforme en valeur absolue (fct abs())
- correction des stocks négatifs qu'on transforme en stock 0 (loc stock quantity < 0 stock quantity = 0)
- vérification des doublons product_id → OK
- vérification des valeurs null + négative dans prix d'achat → OK
- Ajouter test valeur négatifs sur les prix test d'unicité et non nullité sur product_id

Analyses Exploratoires des Données : WEB

DF_Web

	Colonne 1	...	Colonne 29
Ligne 1			
.			
.			
.			
Ligne 1513			



	Colonne 1	...	Colonne 16
Ligne 1			
.			
.			
.			
Ligne 734			

- Table brut des données du site wordpress
- beaucoup de colonnes inutiles : suppression des colonnes qui sont vides, info redondante ou non pertinente (downloadable, post_date, menu_order...)
- Objectif : faire du sku une pk
- sku_manquant + doublons
on garde seulement les sku avec autre colonne non vide.
2 produits sans sku, on recommande de leur attribuer un sku
sku avec un format différent : (13127-1, bon cadeau)
- avertir l'équipe de changer ces sku et d'unifier la nomenclature → Sku doit être en INT avec test de non nullité et d'unicité

Analyses Exploratoires des Données : Liaison

DF_Liaison

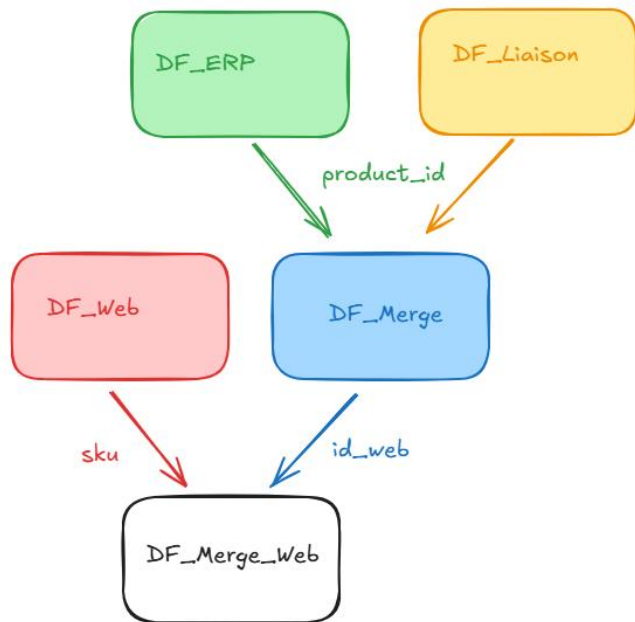
	id_web	product_id
Ligne 4	Null	01
...
Ligne 91	Null	91
...
Ligne 825		



	id_web	product_id
Ligne 1		
...		
Ligne 734		

- Table de correspondance entre erp et web grâce à `id_web` et `product_id`
- 91 `product_id` n'ont pas d'`id_web`
Hypothèse : ces biens ne sont pas vendus sur le site
- Suppression de ces lignes
- On recommande à l'équipe d'attribuer par défaut un `id_web` à chaque produit dans l'erp

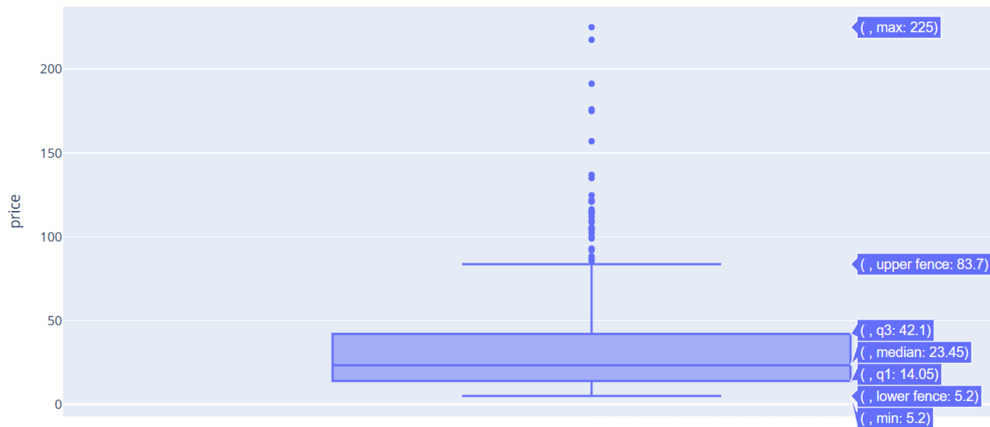
Fusion ou consolidations des données



- Fusion erp + liaison sur un nouveau df : `df_merge` grace à la clé `product_id`.
- On fait une jointure outer avec `indicator = true` pour voir si la correspondance se fait des 2 cotés. On a 91 lignes qui se font en `left_only` [52]. Produit dans l'erp qui n'ont pas d'`id_web`
- Fusion `df_merge` avec `df_web` en faisant correspondre `id_web` et `sku`
- On a des lignes avec un `id_web` mais pas de `sku` qui sont pour la plupart en stock 0. On choisit de les exclure de notre df final

Analyses univariées du prix : visualisation

Répartition des prix grâce à un boxplot (plotly express)



- *Prix des produits situés entre 14 et 42€ avec quelques outliers (>83€)*
- *Quels sont ces outliers? Pourquoi? Est ce qu'il s'agit de valeurs erronées ou aberrantes? Si aberrantes sont elles justifiées?*

Analyses univariées du prix : méthodes statistiques

1ere méthode: par le Z-score

Liste des produits avec Z-score > 3

	post_title	zscore
24	Cognac Frapin VIP XO	5.205993
33	Cognac Frapin Château de Fontpinot 1989 20 Ans...	4.517496
171	Domaine Des Croix Corton Charlemagne Grand Cru...	3.792762
204	David Duband Charmes-Chambertin Grand Cru 2014	6.709816
235	Wemyss Malts Single Cask Scotch Whisky Choc 'n...	3.249211
256	Domaine Weinbach Gewurztraminer Grand Cru Furs...	3.350674
257	Champagne Gosset Célébris Vintage 2007	3.720288
269	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	5.760415
281	Domaine des Comtes Lafon Volnay 1er Cru Champa...	3.212975
282	Domaine des Comtes Lafon Volnay 1er Cru Santen...	3.212975
291	Domaine Clerget Echezeaux Grand Cru En Orveaux...	3.046286
313	Camille Giroud Clos de Vougeot 2016	5.169756
567	Champagne Egly-Ouriet Grand Cru Millésimé 2008	6.981591

- *Z-score ou Z-index : permet de mesurer l'écart de chaque valeur d'un df avec sa moyenne*
- *= (prix - moyenne) / Ecart type*
- *un Z-score élevé permet de déterminer les outliers potentiels*
- *Selon le type de donnée, les outliers sont justifiés ici il s'agit du prix des bouteilles de vin et spiritueux, bien de luxe il est logique d'avoir des prix très élevées*

Analyses univariées du prix : méthodes statistiques

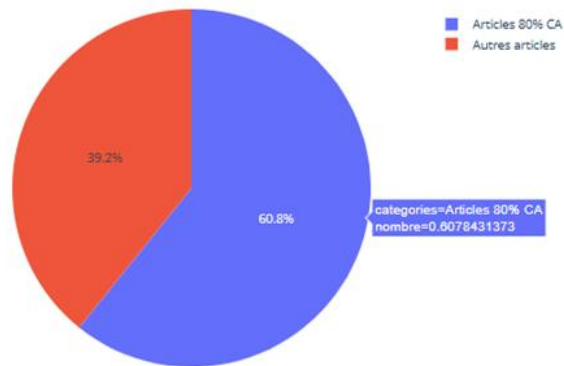
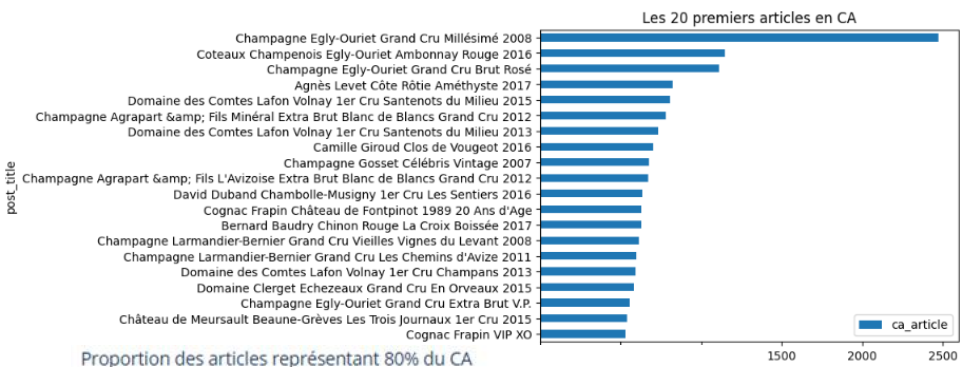
2eme méthode : par l'interval interquartile

Liste des produits avec un prix < Q1 et > Q3

	post_title	zscore
22	Cognac Frapin Château de Fontpinot XO	2.760016
23	Cognac Frapin Cigar Blend	2.596951
24	Cognac Frapin VIP XO	5.205993
33	Cognac Frapin Château de Fontpinot 1989 20 Ans...	4.517496
53	Domaine des Comtes Lafon Volnay 1er Cru Champa...	2.415767
54	Domaine des Comtes Lafon Volnay 1er Cru Santen...	2.633187
56	Zind-Humbrecht Pinot Gris Grand Cru Rangen De ...	2.031658
87	Domaine des Comtes Lafon Volnay 1er Cru Champa...	2.995554
88	Domaine des Comtes Lafon Volnay 1er Cru Santen...	2.633187
89	Château de Meursault Puligny-Montrachet 1er Cr...	2.799876
143	Champagne Larmandier-Bernier Grand Cru Les Che...	1.930195
149	Champagne Agrapart & Fils Minéral Extra Br...	1.973680
150	Champagne Agrapart & Fils L'Avizoise Extra...	2.886844
157	Domaine des Comtes Lafon Volnay 1er Cru Santen...	2.995554
171	Domaine Des Croix Corton Charlemagne Grand Cru...	3.792762
204	David Duband Charmes-Chambertin Grand Cru 2014	6.709816
205	David Duband Chambolle-Musigny 1er Cru Les Sen...	2.654929
233	Wemyss Malts Single Cask Scotch Whisky Chai Ca...	2.959318
234	Wemyss Malts Single Cask Scotch Whisky Chocola...	2.198347
235	Wemyss Malts Single Cask Scotch Whisky Choc 'n...	3.249211
240	Domaine Des Croix Corton Grand Cru Les Grèves ...	2.535348
256	Domaine Weinbach Gewurztraminer Grand Cru Furs...	3.350674
257	Champagne Gosset Célébris Vintage 2007	3.720288
269	Coteaux Champenois Egly-Ouriet Ambonnay Rouge ...	5.760415
281	Domaine des Comtes Lafon Volnay 1er Cru Champa...	3.212975

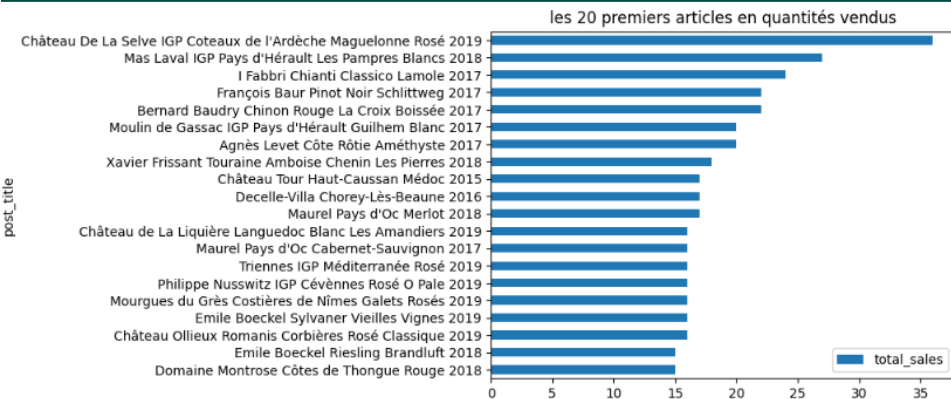
- Calcul des bornes inférieurs et supérieurs
Borne inférieur est négative nous prenons la valeur minimum 5,2€
borne supérieur = $Q3 + ((Q3 - Q1) \times 1,5) = 84\text{€}$
Les outliers sont les prix inférieur à la borne inférieur et supérieur à la borne supérieur

Analyses complémentaires : Vente en CA

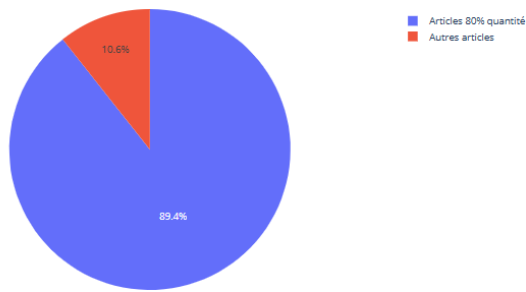


- $CA = \text{somme} (\text{prix vente} \times \text{nb vente}) \rightarrow 143680\text{€}$ sur le mois d'Octobre
- 60% des articles qui font le plus de CA représentent 80% du CA
- Cela ne respecte pas la règle des 20/80

Analyses complémentaires : Vente en quantité

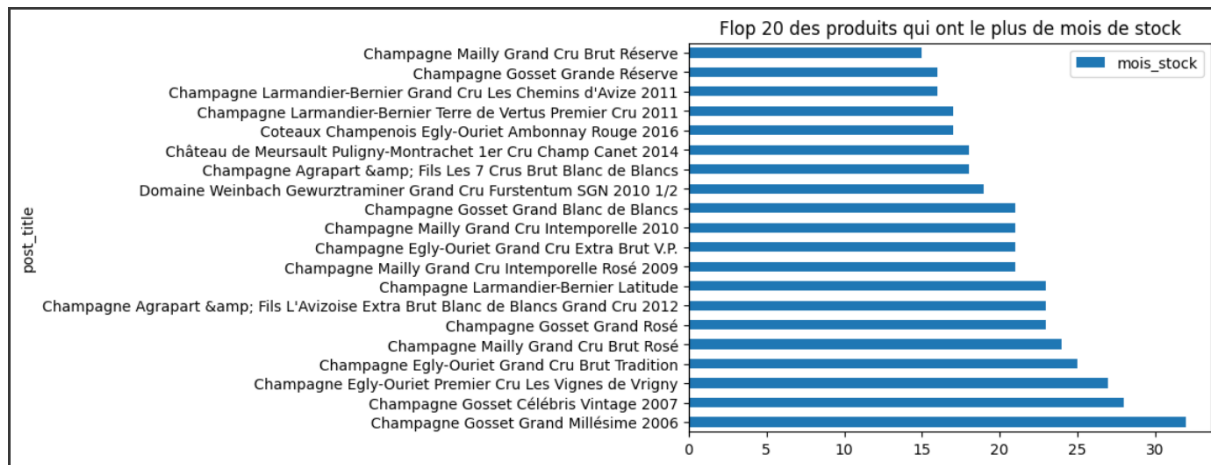


Proportion des articles représentant 80% des quantités



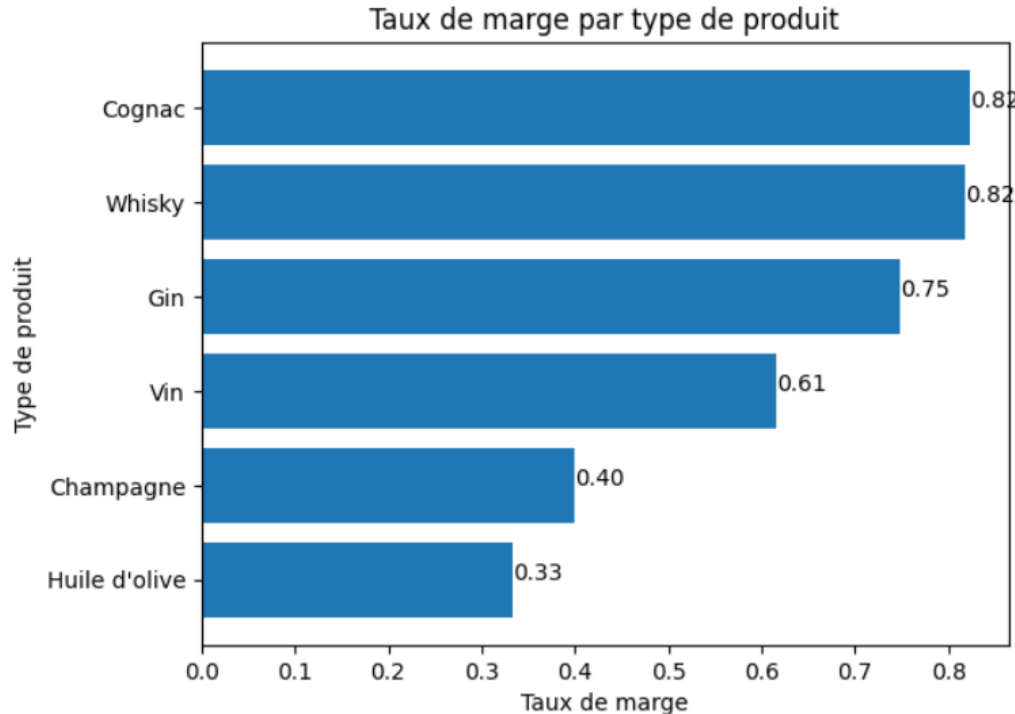
- *89% des articles les plus vendus représente 80% des produits vendus*
- *Cela ne respecte pas non plus la règle des 20/80*

Analyses complémentaires : Stock



- Rotation de stock
Quantité stock / Nb vente (mensuel)
- Nous avons plus de 20 références avec un stock > 1an
- Stock valorisé à 494.638€ pour 16740 produits stockés
- Soit 3,4 mois de stocks

Analyses complémentaires : Taux de marge



- Une réf vente à perte : champagne egly blanc de noir acheté 77,48€ ht et vendu 12,65 ttc faire remonter pour vérifier si erreur
- Type de produit qui présente une bonne et une mauvaise marge

Analyses complémentaires : Corrélations



- *Corrélation négative modéré entre prix et vente : + prix augmentent et plus les ventes diminuent*
- *Corrélation positive modéré entre quantité et vente : bonne gestion du stock, plus le produit se vend et plus on anticipe un stock élevé de sécurité afin d'éviter le stock 0*
- *Corrélation négative faible entre prix et quantité*

Actions pour la suite : recommandation technique

ERP :

- Définir *product_id* en INT
- Ajouter test valeur négatifs sur les prix + test d'unicité et non nullité sur *product_id*

WEB:

- Définir *sku* en INT
- Corriger les *sku* manquants et mal renseignés
- Ajouter test d'unicité et non nullité sur *SKU*
- colonne *taxe* : Boolean false or true

DF_Final :

- Règle si *ONSALEWEB* = 1 alors *SKU* ne doit pas être null
- Définir règle stock ≤ 0 alors Outofstock (Suggestion mettre alerting en place si stock < x quantité)

Documenter les tables, définir la nomenclature

Point sur les compétences apprises

- *Qu'est-ce qui s'est bien passé pour vous dans ce travail de nettoyage ? La compréhension des dataframes et les enjeux métiers*
- *Qu'est-ce que vous avez trouvé le plus difficile ? nettoyage des données. Prendre les bonnes décisions en fonction des erreurs potentiels*
- *Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ? Les fonctions pythons (syntaxe) et le processus de nettoyage. Il s'agit de nouvelles compétences sur lesquelles j'ai besoin de m'entraîner davantage.*