

Optimisez la gestion des données d'une boutique

Stéphane Srsa
FAO



Contexte



J'ai intégré la société **bottleneck** pour réaliser une mission de Data Analyst.
Bottleneck est un marchand de vin très prestigieux

Situation actuelle

- Utilisation d'un ERP non relié au site de vente en ligne
- Une table provenant de l'outil CMS (vente en ligne)
- Une table rapprochant les références ERP et WEB
- Gestion des stocks complexe
- Analyse des ventes en ligne réduite

Méthodologie de l'analyse

Information concernant les fichiers utilisés pour l'étude

ERP : référence produit – prix – état des stock

WEB : référence produit en ligne – Nbr ventes – description...

Liaison : références ERP – références WEB

Analyse

- Analyse exploratoire des différents fichiers.
- Mise en cohérence des colonnes, nettoyage des données, vérification des types de données et de l'unicité de la clé choisie.
- Création d'une table unique après rapprochement des 3 tables.
- Réponses aux demandes de Laurent

Analyse des fichiers fournis

Fichier ERP

- ERP contient 825 lignes et 5 colonnes
- Les Dtypes sont cohérents avec les données
- Il n'y a pas de doublons
- Choix de la clé => 'product_id'
- Vérification de l'unicité de la clé => Ok
- D'après 'onsale_web', 108 références ne sont pas proposées en ligne

	product_id	onsale_web	price	stock_quantity	stock_status
820	7203	0	45.0	30	instock
821	7204	0	45.0	9	instock
822	7247	1	54.8	23	instock
823	7329	0	26.5	14	instock
824	7338	1	16.3	45	instock

Analyse des fichiers fournis

Fichier Liaison

- Liaison contient 825 lignes et 2 colonnes
- Le Dtypes 'id_web' est changé en int64
- Il n'y a pas de doublons
- Choix de la clé => 'product_id'
- Vérification de l'unicité de la clé => Ok
- Après traitement, 'id_web' comporte 91 références non renseignées et 3 références incohérentes

	product_id	id_web
820	7203	NaN
821	7204	NaN
822	7247	13127-1
823	7329	14680-1
824	7338	16230

product_id int64
id_web int64
dtype: object

product_id	id_web
4053	13127
4757	14680
4954	bon-cadeau-25-euros
7247	13127-1
7329	14680-1

Analyse des fichiers fournis

Fichier WEB – étapes analyse/traitement

Voir le
Notebook
pour plus
de détails

- WEB contient 1513 lignes et 28 colonnes
- **Suppression des 11 colonnes ne contenant aucune données**
- Renommer 'sku' en 'id_web' est changer Dtype en int64
- **Suppression des 83 lignes ne contenant aucune données**
- Vérification et enregistrement des 2 valeurs nulles restantes sur 'id_web'
- **Vérification et enregistrement des 4 valeurs incohérentes sur 'id_web'**
- Détection et suppression des doublons
- **Création d'une table 'Bons Cadeaux' et suppression de la tables des références 'vins'.**
- Changer les Dtypes objet en datetime pour les dates
- **Choix de la clé => 'id_web'**
- Vérification de l'unicité de la clé => Ok
- **Après traitement, WEB comporte 712 références renseignées.**

Jointure des tables

Tables ERP et Liaison => liaisonerp

- Utilisation de la condition 'outer' pour la jointure
- Toutes les jointures sont de nature 'Both'
- Il y a 94 lignes avec un 'id_web' non renseigné

Tables WEB et liaisonerp => datawine

- Utilisation de la condition 'outer' pour la jointure
- Il y a 113 jointures différentes de 'Both'
- Il y a 94 lignes avec un 'id_web' non renseigné
- Il y a 19 lignes à vérifier et à conserver dans un DF

Table 'Bon_cadeaux' : 1 ligne

id_web	total_sales	tax_status	post_author	post_title
0	10.0	taxable	1.0	Bon cadeau de 25€

Table 'datawine' : 712 lignes

	product_id	id_web	onsale_web	price	stock_quantity	stock_status	total_sales	tax_status
711	7338	16230	1	16.3	45	instock	0.0	taxable

Fichiers créés et non utilisés pour la suite de l'analyse

Dataframes enregistrés pour investigation si besoin

DF concernant les valeurs incohérentes du fichier 'web' => web_inco

DF concernant les 94 id_web non renseignés => idweb0

DF concernant les 19 références web non renseignées => datawine_valeurs_manquantes

DF concernant les 2 références web renseignées sans 'id_web' => web_Ref_idweb0

id_web	total_sales	tax_status	post_author	post_title
bon-cadeau-25-euros	10.0	taxable	1.0	Bon cadeau de 25€
13127-1	0.0	taxable	2.0	Clos du Mont-Olivet Châteauneuf-du-Pape 2007
bon-cadeau-25-euros	10.0	NaN	1.0	Bon cadeau de 25€
13127-1	0.0	NaN	2.0	Clos du Mont-Olivet Châteauneuf-du-Pape 2007

	product_id	id_web	onsale_web	price	stock_quantity	stock_status	total_sales	tax_status	post_author
93	7329	0	0	26.5	14	instock	NaN	NaN	NaN

	product_id	id_web	onsale_web	price	stock_quantity	stock_status	total_sales	tax_status	post_author
18	6100	15529	0	12.9	0	outofstock	NaN	NaN	NaN

	id_web	total_sales	tax_status	post_author	post_title
470	NaN	0.0	taxable	2.0	Pierre Jean Villa Côte Rôtie Fongean 2017
471	NaN	0.0	taxable	2.0	Pierre Jean Villa Condrieu Jardin Suspendu 2018

Calcul du chiffre d'affaires par produit

Process

Création d'une colonne 'Ca_produit' contenant le produit du prix par la quantité vendue

	product_id	id_web	price	total_sales	Ca_produit	stock_quantity	post_title	post_excerpt
194	4334	7818	49.0	96.0	4704.0	0	Champagne Gosset Grand Blanc de Blancs	La bulle fine et presente se dirige vers la su...
71	4144	1662	49.0	87.0	4263.0	11	Champagne Gosset Grand Rosé	Le nez exprime sans détour la fraise, dans sa ...
218	4402	3510	176.0	13.0	2288.0	8	Cognac Frapin VIP XO	La cuvée VIP XO à été enrichie d'eaux-de-vie t...

Chiffre d'affaires global réalisé en ligne

Process

Somme des Ca des produits vendus en ligne
Ajout du Ca réalisé avec les bons cadeaux

	Chiffre d'Affaires	Nbr Références
Chiffre d'affaires bouteilles sur le web	70,318.60	712
Chiffre d'affaires 'bon cadeaux' sur le web	250.00	1
Chiffre d'affaires global sur le web	70,568.60	713

Analyse sur la variable 'prix'

Process

- Détection des Outliers avec la méthode des interquartiles
- Calcul des interquartiles
- Création d'un Dataframe reprenant les Outliers repérés
- Visualisation graphique des Outliers

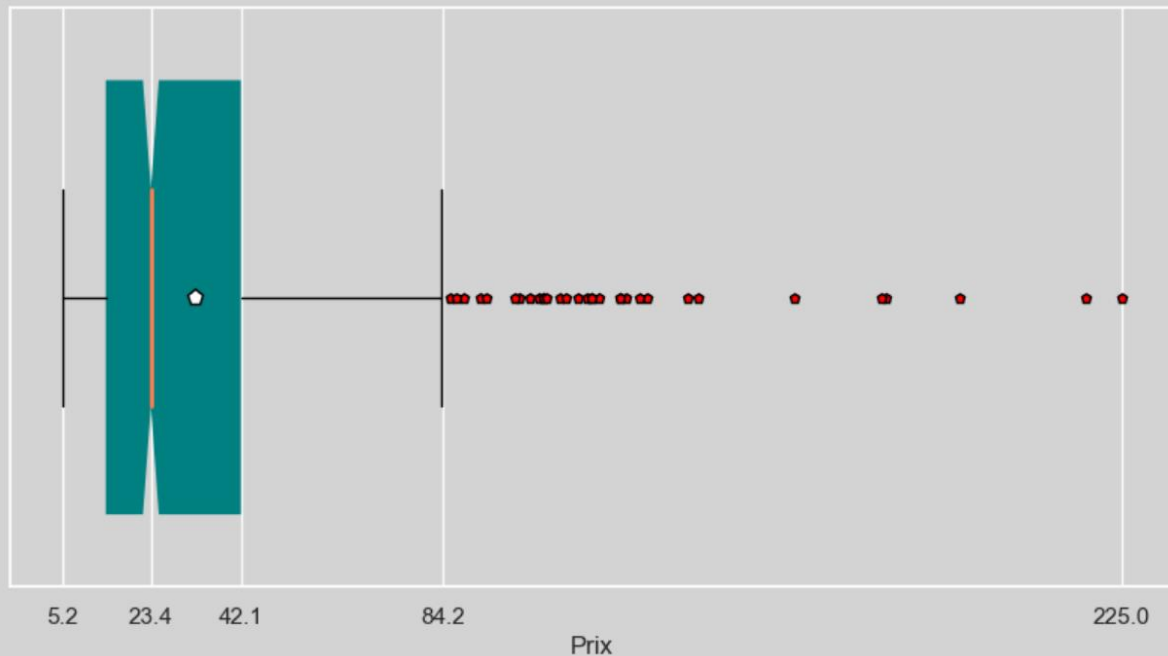
	Valeur	product_id	id_web	price	total_sales	Ca_produit	stock_quantity	post_title
IQR	28.04	4352	15940	225.0	5.0	1125.0	0	Champagne Egly-Ouriet Grand Cru Millésimé 2008
Q1	14.09							
Q3	42.12	5001	14581	217.5	0.0	0.0	20	David Duband Charmes- Chambertin Grand Cru 2014
Limite basse	-27.97							
Limite haute	84.18	5892	14983	191.3	3.0	573.9	10	Coteaux Champenois Egly- Ouriet Ambonnay Rouge ...

Visualisation des Outliers

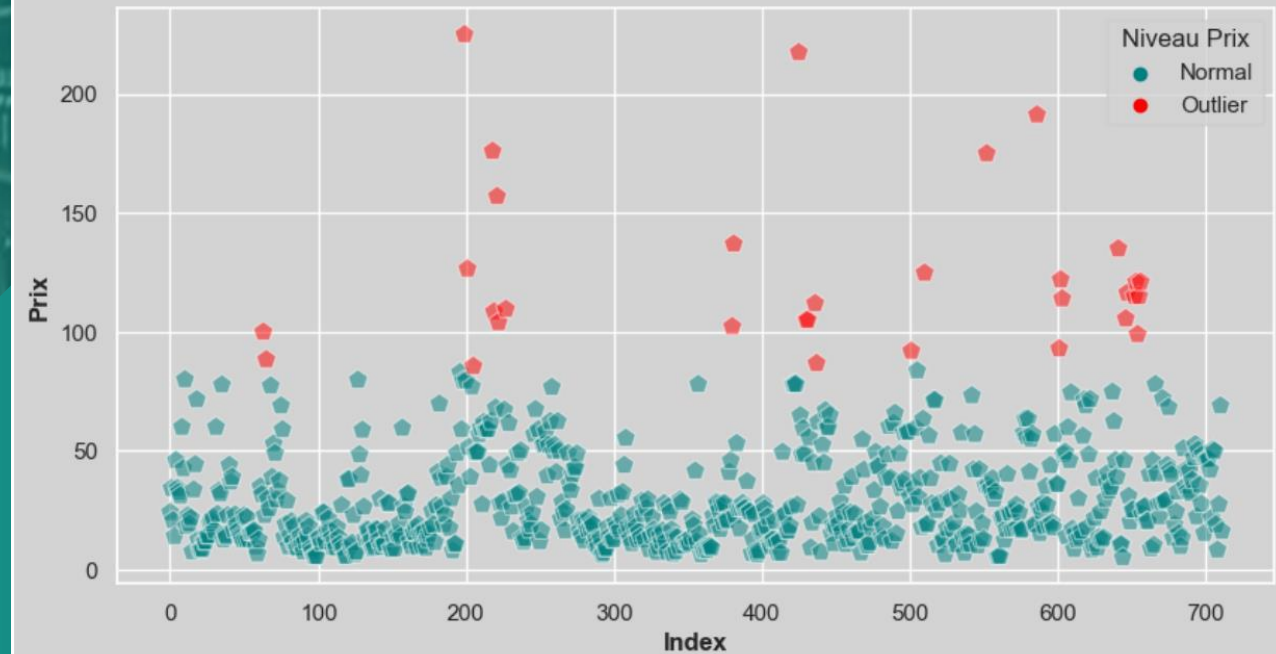
Observation

Il y a 32 Outliers à analyser

Affichage des Outliers - Boxplot



Affichage des Outlier - Scatterplot



Vérification des 32 références dites 'outliers'

Process

- Création d'une liste de liens hypertextes pour la recherche des références suivant leurs intitulées.

63

<https://www.google.com/search?q=ZindHumbrecht+Riesling+Grand+Cru+Rangen+De+Thann+Clos+Saint-Urbain+2017&start>

65

<https://www.google.com/search?q=ZindHumbrecht+Pinot+Gris+Grand+Cru+Rangen+De+Thann+Clos+Saint-Urbain+2012&start>

199 <https://www.google.com/search?q=Champagne+EglyOuriet+Grand+Cru+Millésimé+2008&start>

201 <https://www.google.com/search?q=Champagne+EglyOuriet+Grand+Cru+Blanc+de+Noirs&start>

218 <https://www.google.com/search?q=Cognac+Frapin+VIP+XO&start>

219 <https://www.google.com/search?q=Cognac+Frapin+Château+de+Fontpinot+XO&start>

Conclusion

Tous les prix sont cohérents, y compris les 'Outliers'.
Les liens hypertextes le confirment

Analyse des stocks sur les 10% des références les plus vendues

Méthode

- Création d'un Dataframe intégrant les meilleurs ventes ainsi que l'alerte stock et le besoin de commander ou pas.
- Alerte basée sur 10% du CA réalisé. (Règle à définir avec Laurent)
- Il y a 384 références vendues au moins une fois, nous retenons les 40 meilleurs ventes pour l'analyse

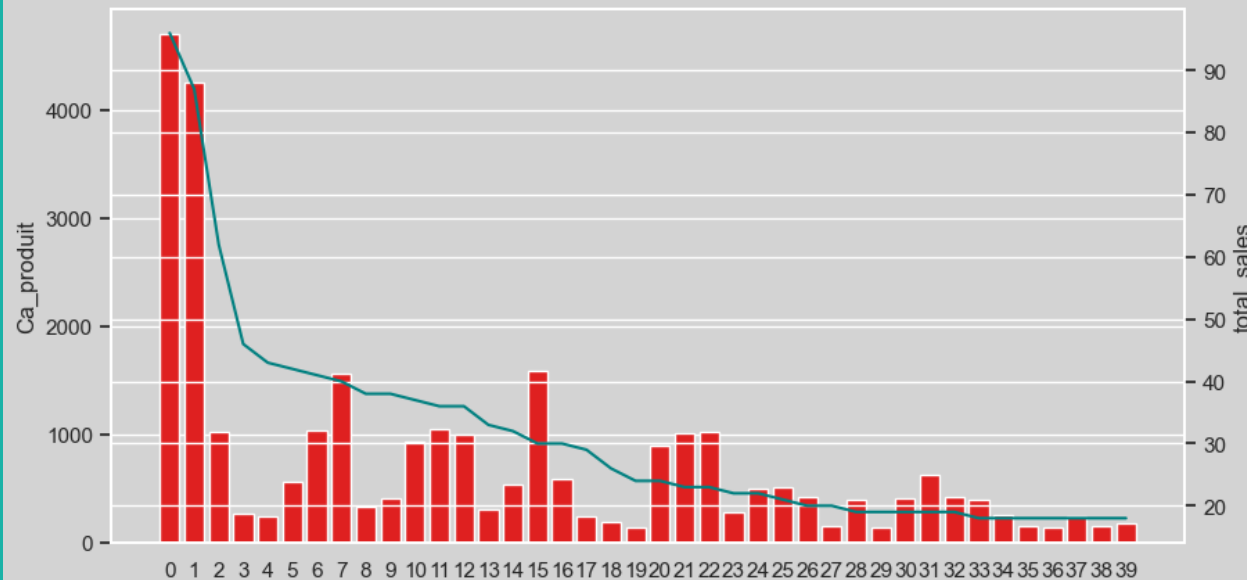
product_id	id_web	price	total_sales	Ca_produit	Stock quantity	post_title	alerte stock	Alerte
4334	7818	49.0	96.0	4704.0	0	Champagne Gosset Grand Blanc de Blancs	470.0	A commander
4144	1662	49.0	87.0	4263.0	11	Champagne Gosset Grand Rosé	426.0	A commander
4068	16416	16.6	62.0	1029.2	157	Gilles Robin Crozes-Hermitage Rouge Papillon 2019	103.0	Stock ok

Analyse des stocks sur les 10% des références les plus vendues

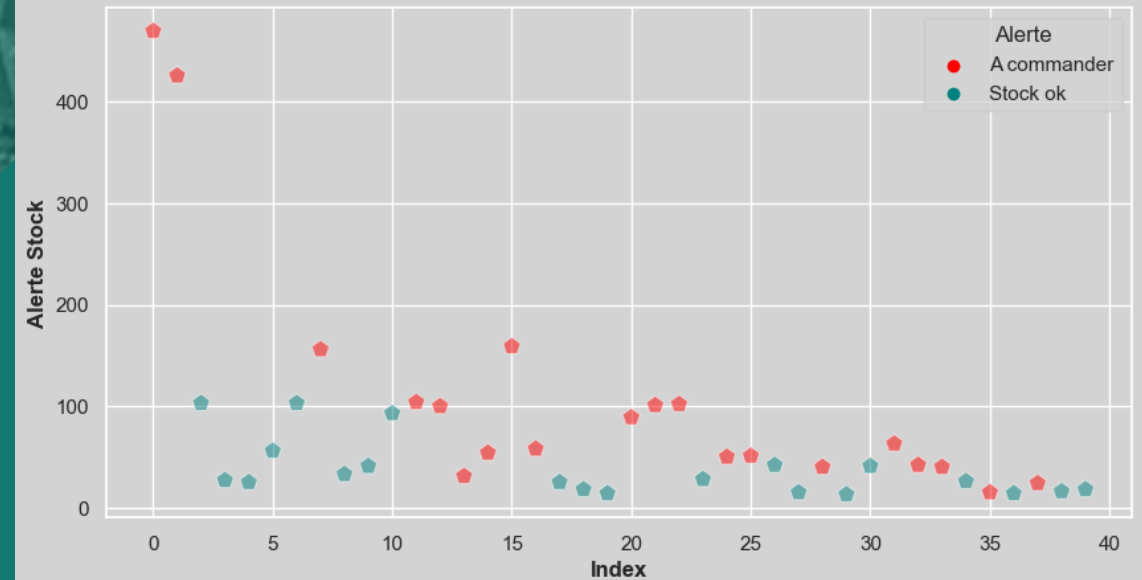
Visualisation

- Graphique du CA et quantité vendue par référence
- Graphique de l'état des Stocks par référence

CA et quantité des références les plus vendues



Etat des stocks des meilleurs ventes



Analyse des stocks sur les 10% des références les plus vendues

Mise en avant des produits à commander

- Sur les 40 références les plus vendues, il y a 20 produits à commander rapidement

	product_id	id_web	price	total_sales	Ca_produit	Stock quantity	post_title	alerte stock	Alerte
0	4334	7818	49.0	96.0	4704.0	0	Champagne Gosset Grand Blanc de Blancs	470.0	A commander
1	4144	1662	49.0	87.0	4263.0	11	Champagne Gosset Grand Rosé	426.0	A commander
17	4731	8344	22.0	18.0	396.0	7	Domaine Muré Crémant d'Alsace Rosé	40.0	A commander
18	6042	16159	8.5	18.0	153.0	0	Domaine La Croix Belle Côtes de Thongue Rosé Les Grillons 2019	15.0	A commander
19	5439	15649	13.2	18.0	237.6	0	La Préceptorie Maury Sec Rouge Copain Comme Cochon 2018	24.0	A commander