

Optimisez la gestion du stock d'une boutique en nettoyant ses données

Frédéric NALLET

Business Intelligence Analyst

19/12/2023

Analyses Exploratoires des Données

Datasets :

	Fichiers			
	web	liaison	erp	caracteristiques_vins
Format fichier	Fichier xls	Fichier xls	Fichier xls	Fichier csv
Observations	1513	825	825	611
Caractéristique	28	2	5	13
Anomalies :				
Données incohérentes	1			
Typologie erronée	2			
Valeurs manquantes	85	91	0	118
Doublons	716	90	0	0

Analyses Exploratoires des Données

Traitement réalisés :

Nettoyage des données :

- Suppression des colonnes inutiles
- Analyse des données
- Suppression des doublons
- Suppression des observations incomplètes

Features engineering :

- Comparaison des données existantes avec des données recalculées.
- Création de nouvelles variables.

Analyses Exploratoires des Données

- Analyser les formats de fichiers et leur données
- Analyser les données incohérentes
- Analyser les données qui peuvent être recalculable
- Analyser les colonnes aux données manquantes

Fusion ou consolidations des données

La Jonction entre les fichiers « df_erp » et « df_liaison » :

Clés communes : « product_id »

Jointure : « Outer »

La Jonction entre les fichiers « df_merge » et « df_web » :

Clés : « id_web » et « sku »

Jointure : « Outer »

La Jonction entre les fichiers « df_merge » et « df_caractéristiques » :

Clés communes : « post_name »

Jointure : « Left »

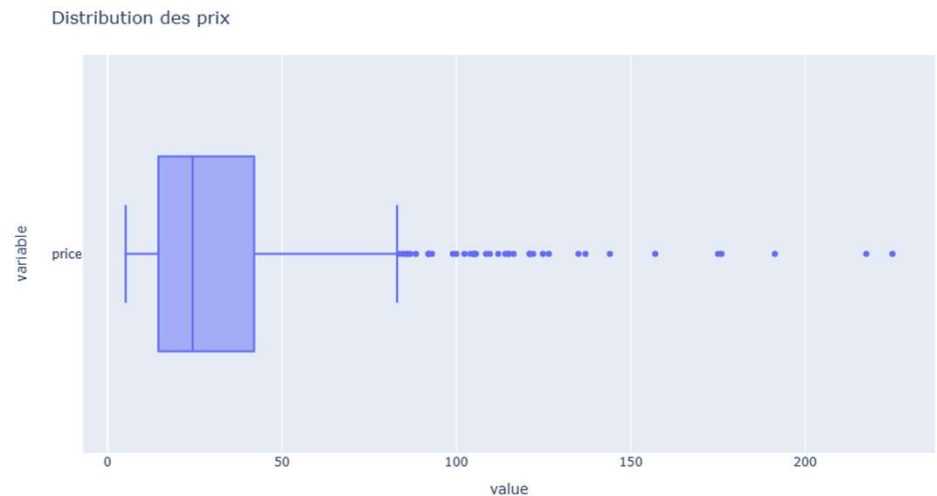
Fusion ou consolidations des données

- Déterminer les attributs des clés
- Attributs en doublon : « sku » et « id_web »
- Identifier les données non fusionnées

Analyses univariées du prix

Méthodes statistiques employées :

- La moyenne
- L'écart-type
- Le Z-score
- Le seuil de prix



Le seuil de prix n'est pas représentatif car certains produits sont luxueux et donc éloignés de la moyenne.

Z-score = 3 afin de reprendre le groupe de données des produits luxueux

Analyses univariées du CA

Loi de pareto :

80 % du chiffre d'affaire :

- 130 articles
- 18% des articles

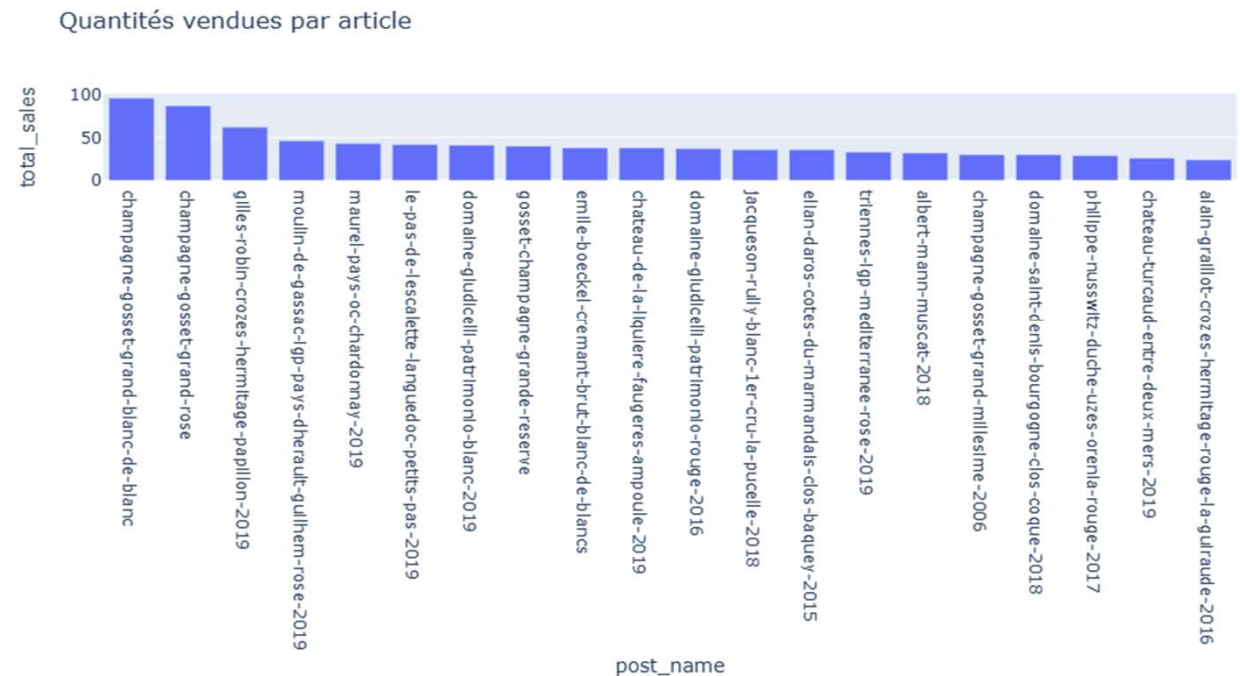


Analyses univariées du CA

Loi de pareto :

80 % des produits vendus :

- 150 articles
- 21% des articles



Actions pour la suite

- Analyser les produits en fonction de leur gamme de prix
- Analyser les stocks pour éviter la pénurie
- Mettre en place les dates de sorties

Point sur les compétences apprises

Mes compétences acquises :

- Lecture des fichiers et les jonctions
- Identifier et suppression des doublons

Mes compétences à améliorer :

- Correction des données incohérentes ou manquantes

Mes compétences à compléter :

- Manipuler les statistiques