

[Bottleneck]

[Brochart Pascal]

[Data Analyst]

[26/03/2024]

Contexte du projet

- *L'objectif de cette étude est de faire un rapprochement et une analyse sur les différents produits de Bottleneck, un marchand de vin prestigieux*
- *Un nettoyage des données est nécessaire pour la liaison des extractions de l'ERP et du site Web*
- *L'analyse doit porter notamment sur la gestion des stocks et la corrélation entre certaines données quantitatives*
- *Les tops références, le 20/80, chiffres d'affaires et détection de valeurs aberrantes seront également abordés*

Analyses Exploratoires des Données

- *Le fichier erp contient les produits, la quantité de stocks, le prix d'achats et de vente*
- *Le fichier web contient les types et la description des produits, le nombre de ventes et la gestion des fiches produits sur le site web*
- *Le fichier liaison contient les liens permettant de relier les produits de l'ERP avec les produits du site Web*
- *Le fichier web contient beaucoup de colonnes contrairement aux fichiers erp et liaison, les données sont fournies sous format Excel*

Détection des anomalies

- *Clé 'sku' avec des valeurs en chaine de caractères:*
 - *13127-1*
 - *bon-cadeau-25-euros*
- *4 produits avec un prix d'achat supérieur au prix de vente*
 - *Product_id 7196, 4864, 4355 et 6324*
- *2 Valeurs erronées de stock_status*
 - *Product_id 4039 (instock au lieu de outofstock) et 4885 (outofstock au lieu de instock)*
- *Doublement des lignes web avec la clé 'sku' (une avec post_type 'attachment' et une avec 'product') on ne garde que cette dernière*
- *2 valeurs négatives de quantité de stocks (-10 et -1)*
- *3 valeurs négatives de prix de vente (-20€, -9,1€ et -8€)*

Analyses Exploratoires des Données

- *Des valeurs de prix et des quantités de stocks sont négatives*
- *Traitement réalisés*
 - *Beaucoup de colonnes dans le jeu de données web ont été supprimées car inutiles ou non exploitables*
 - *Des valeurs incohérentes ont été détectées et corrigées notamment pour la colonne 'stock_status' qui obéit à une règle fonctionnelle simple*
- *Remarques éventuelles, pièges ou difficultés rencontrées*
 - *Vérification des doublons sur les produits*

Fusion ou consolidations des données

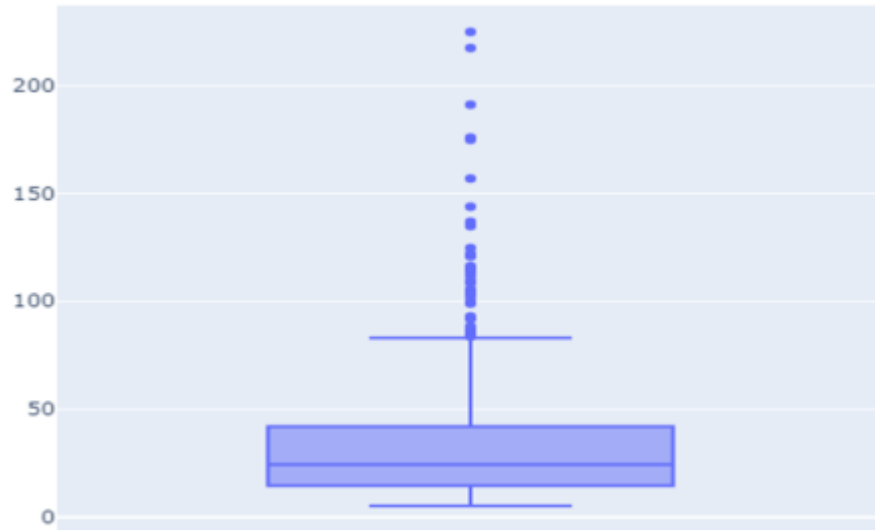
- *La fusion des données erp avec web se fait par l'intermédiaire du fichier liaison*
- *L'ERP est joint avec la clé 'product_id' qui permet d'identifier de manière unique le produit*
- *La mapping avec les données web est fait avec la clé 'sku' ('id_web' dans liaison qui sera renommé pour l'occasion)*
- *Les clés seront donc 'product_id' et 'sku'*
- *Il conviendra d'éliminer les doublons dans la partie web et ne garder que les ligne avec un post_type de valeur 'product'*

Analyses univariées du prix

- *Méthode d'identification par le Z-index*
 - *Ecart type du prix est de 26€*
 - *Le seuil du prix Z-score > 3 est 112€*
- *Méthode d'identification par l'intervalle interquartile*
 - *Le seuil du prix abbérant est 84€*
 - *Correspond à 36 articles au dessus du seuil*

Ces méthodes visent à déterminer le seuil entre les valeurs jugées normales et celles dites aberrantes

Analyses univariées du prix

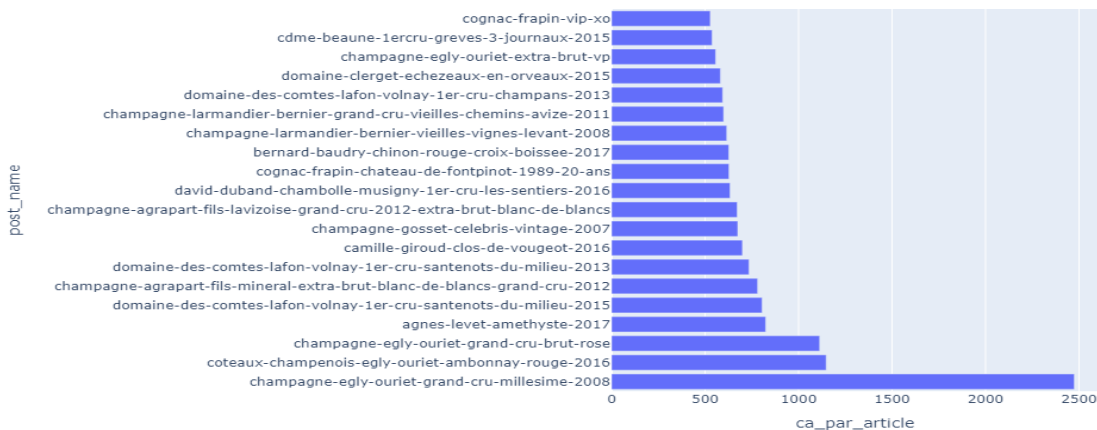


- *Les bornes supérieures et inférieures définissent la limite des valeurs normales, la médiane se situe vers 25€ et les valeurs aberrantes au dessus de 85€*

- *On obtient le Chiffre d'affaire en multipliant le nombre de ventes par le prix à l'unité
Celui-ci s'établit à 143680€*
- *Les 80% du chiffre d'affaire ou des ventes en quantité sont obtenus par une somme cumulative des valeurs triées dans l'ordre décroissant*
 - *434 articles et 53% du catalogue pour le CA*
 - *424 articles et 51% du catalogue pour les quantités*
- *La valorisation du stock est la multiplication des quantités en stocks par le prix de vente soit 532119€*

Analyses complémentaires CA par article

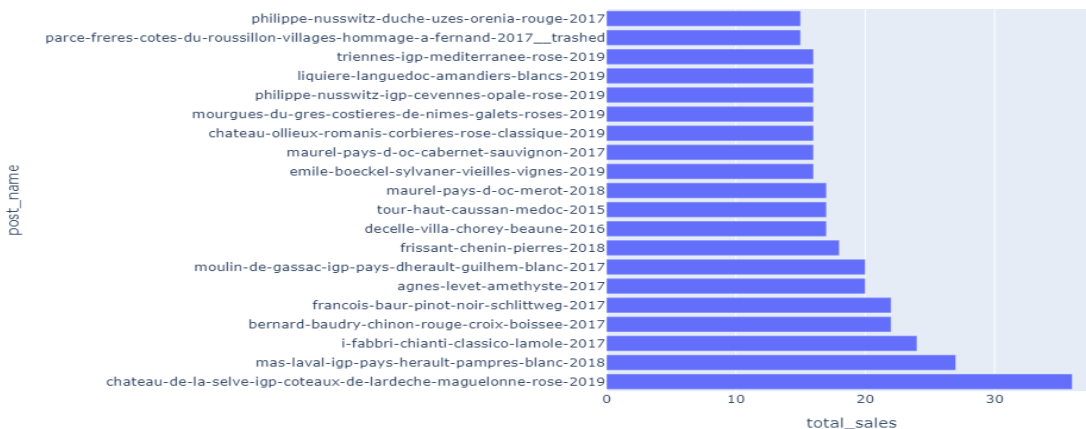
Graphique en barre des 20 premiers articles



- Les grands crus renommés figurent dans ce graphique

Analyses complémentaires CA par quantités

Graphique en barre des 20 premiers articles



- Les vins plus accessibles figurent dans ce graphique

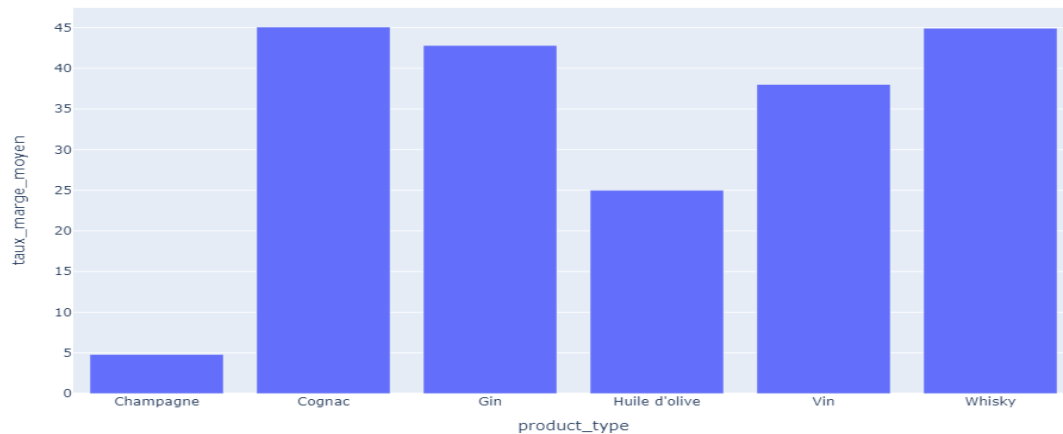
- *Le taux de marge est obtenu par la division de la marge brut et le prix de vente HT
La marge brut est la soustraction entre le prix de vente HT et le prix d'achat*
- *Une corrélation positive forte proche de 1 permet de mettre en évidence que lorsqu'une variable augmente l'autre aura tendance à augmenter également*
- *A l'inverse une corrélation négative forte proche de -1 permet de mettre en évidence la diminution d'une variable lorsque l'autre augmente*

Analyses complémentaires

Taux de marge par produit



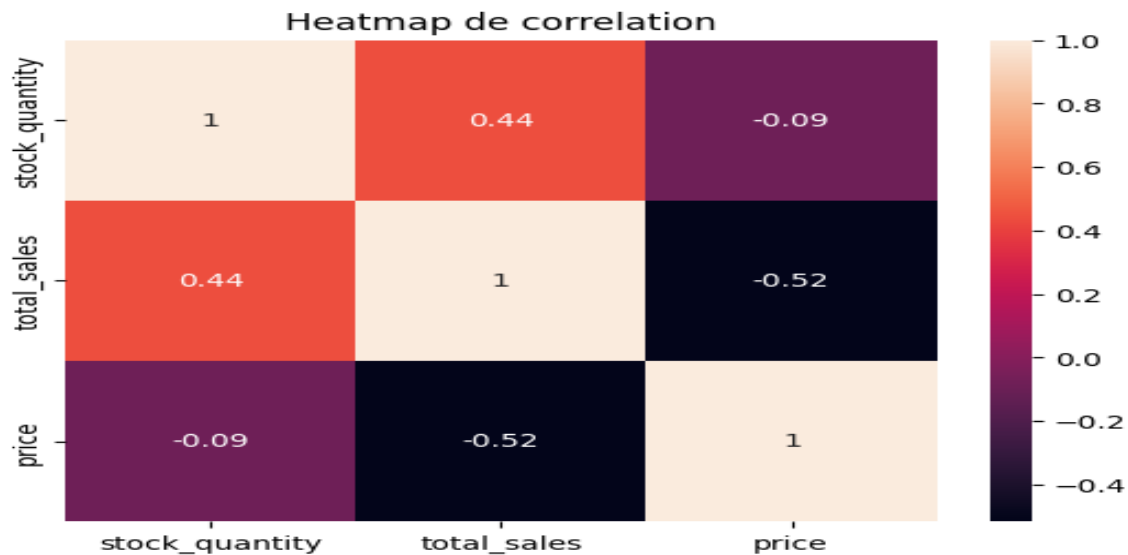
Graphique du taux de marge par type de produit



- Le champagne a le taux de marge le plus bas

Analyses complémentaires

Corrélations



- Le nombre de ventes diminue lorsque le prix augmente avec une corrélation négative de -0,52

Synthèse du projet

- *Le nettoyage et la correction de certaines données ont permis le rapprochement des données de l'ERP avec le site Web*
- *La détection des données outliers a été réalisée grâce aux analyses d'identifications par le Z-index et l'intervalle interquartile*
- *La relation de certaines valeurs a pu être mise en évidence avec la création d'une matrice de corrélation et d'une heatmap*
- *Le taux de marge a été calculé pour chacun des produits en fonction de son prix d'achat*

Actions pour la suite

- *Certaines données méritent une analyse plus pertinente et pourront être transmises à d'autres membres de l'équipe*
- *Le nettoyage des données peut être encore amélioré en coordination avec le reste de l'équipe sur les éventuelles fonctionnalités à venir ou restant à développer*
- *Des contrôles de cohérence peuvent être mis en place afin d'éviter l'enregistrement de données non conformes*
- *S'assurer que le projet respecte le RGPD*