

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

- Mission : Rapprochement des données entre ERP et boutique en ligne
- Objectif : Améliorer la gestion des stocks et l'analyse des ventes en ligne

Tahir Tahirzade
[Business Intelligence Analyst]
04/08/2023

— Analyses Exploratoires des Données

| | | | |
|---|--|---|---|
| <u>Erp.xls</u> | <u>Caractéristique:</u> <ul style="list-style-type: none">• <u>Nbr variables</u>: 5 colonnes• <u>Clés uniques</u>: 825• <u>Nbr observation</u> : 825 lignes | <u>Traitement réalisés:</u> <ul style="list-style-type: none">• Analyse variable 'stock_status': erreur identifiée et corrigée (ligne 443)• Variables 'stock_status' et 'stock_status_2' supprimées car redondance avec 'stock_quantity' | <u>Remarques éventuelles:</u> <ul style="list-style-type: none">• A première vue, la variable 'stock_status' paraissait correctement renseignée mais une erreur s'y était glissée |
| <u>Web.xlsx</u> | <u>Caractéristique:</u> <ul style="list-style-type: none">• <u>Nbr variables</u>: 28 colonnes• <u>Clés uniques</u>: 714• <u>Nbr observation</u> : 1513 lignes | <u>Traitement réalisés:</u> <ul style="list-style-type: none">• Suppression de 20 colonnes dont les informations n'étaient pas nécessaires à l'analyse• Identifications et suppression des doublons de type 'attachement'• Identification et suppression des lignes sans codes article ('sku') | <u>Remarques éventuelles:</u> <ul style="list-style-type: none">• Difficulté à trier les 28 colonnes pour conserver uniquement les variables pertinentes• Difficulté dans le choix d'écarter ou non les 2 observations 'sans sku' |
| <u>Liaison.xls</u> | <u>Caractéristique:</u> <ul style="list-style-type: none">• <u>Nbr variables</u>: 2 colonnes• <u>Clés uniques</u>: 734• <u>Nbr observation</u> : 825 lignes | <u>Traitement réalisés:</u> <ul style="list-style-type: none">• Pas de nettoyage réalisée à ce stade | <u>Remarques éventuelles:</u> <ul style="list-style-type: none">• 91 observations n'ont pas de correspondances entre 'product_id' et 'id_web' |
| <u>Caractéristiques vins.csv</u> | <u>Caractéristique:</u> <ul style="list-style-type: none">• <u>Nbr variables</u>: 13 colonnes• <u>Clés uniques</u>: 611• <u>Nbr observation</u>: 611 lignes | <u>Traitement réalisés:</u> <ul style="list-style-type: none">• Pas de nettoyage réalisée à ce stade | <u>Remarques éventuelles:</u> <ul style="list-style-type: none">• Erreur de lecture du fichier csv (utilisation 'chardet_detect' pour déterminer encodage)• 118 observations ont des données manquantes |

Fusion ou consolidations des données

| Jonction du fichier df_erp et df_liaison | Jonction du fichier df_merge et df_web | Jonction du fichier df_merge et df_caracteristiques |
|---|--|--|
| <u>Clés utilisées :</u> <ul style="list-style-type: none">• 'product_id' dans les 2 dataframes | <u>Clés utilisées:</u> <ul style="list-style-type: none">• 'id_web' dans le dataframe 'df_merge'• 'sku' dans le dataframe 'id_web' | <u>Clés utilisées:</u> <ul style="list-style-type: none">• 'post_name' dans les 2 dataframes |
| <u>Choix des attributs:</u> <ul style="list-style-type: none">• Merge(on=product_id,how='outer', indicator=true) | <u>Choix des attributs:</u> <ul style="list-style-type: none">• Merge(df_web,left_on='id_web',right_on='sku',how='left',indicator=true) | <u>Choix des attributs:</u> <ul style="list-style-type: none">• df_merge=df_merge.merge(df_caracteristique, on=post_name, how='left', indicator=true) |
| <u>Vigilances particulières au cours du traitements:</u> <ul style="list-style-type: none">• S'assurer de la correspondance des observations lors de la jonction | <u>Vigilances particulières au cours du traitements:</u> <ul style="list-style-type: none">• S'assurer de la correspondance des observations lors de la jonction | <u>Vigilances particulières au du traitements:</u> <ul style="list-style-type: none">• S'assurer de la correspondance des observations lors de la jonction |
| <u>Difficultés ou pièges rencontrés:</u> <ul style="list-style-type: none">• Aucunes difficultés | <u>Difficultés ou pièges rencontrés:</u> <ul style="list-style-type: none">• 111 lignes ont une correspondance uniquement à gauche car seulement 714 'sku' renseignées dans le dataframe 'df_web' | <u>Difficultés ou pièges rencontrés:</u> <ul style="list-style-type: none">• 214 lignes ont une correspondance uniquement à gauche car seulement 611 observations dans le dataframe 'df_caractéristiques' |

Analyses univariées du prix

Méthodes statistiques employés

Z-score:

- Calcul de la moyenne du prix
- Calcul de l'écart-type du prix
- Calcul du Z-Score

Intervalle interquartile :

- Calcul de l'intervalle interquartile
- Analyse des outliers

- Création d'une Boîte à moustache avec Plotly Express

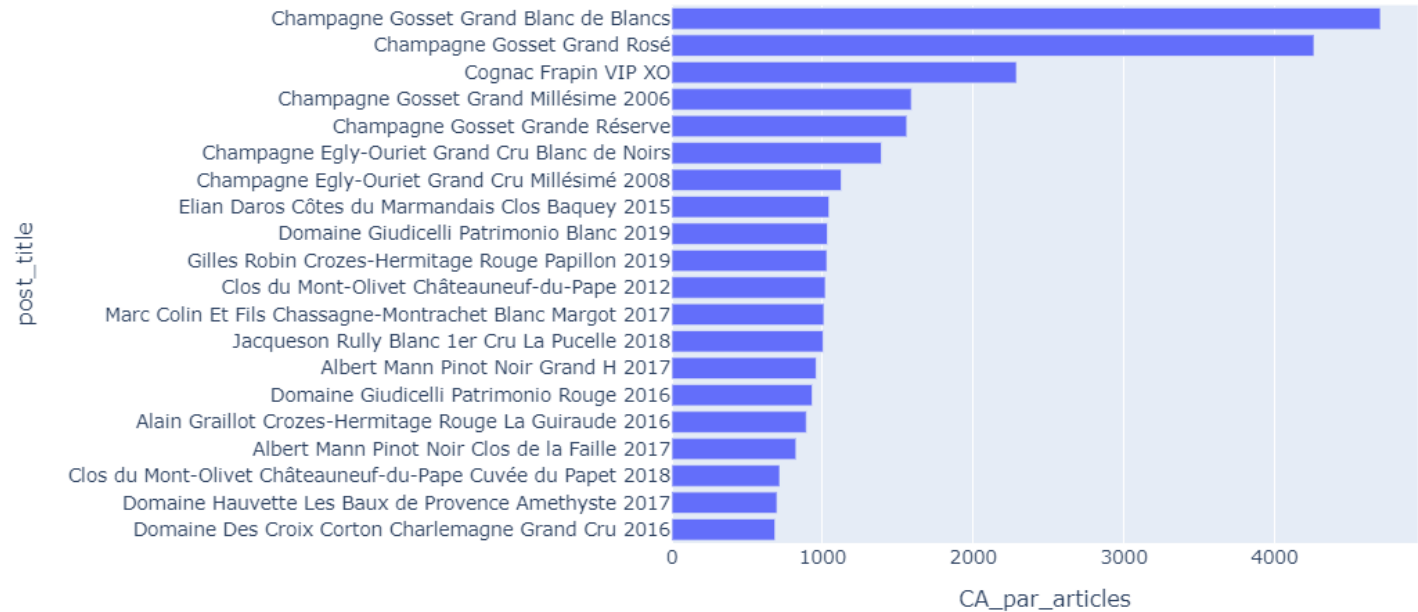
Boxplot price



Analyses univariées du CA

| Méthodes statistiques employés: | Graphique avec commentaire des résultats: | Limites éventuelles de l'analyse : |
|--|---|--|
| Clacul Total CA :70568.6€ | 6 des 7 articles réalisant le plus de CA sont des champagne | Certain produits pouvant potentiellement fausser l'analyse |
| Calcul du20/80 en CA: 351 articles=80% | | |

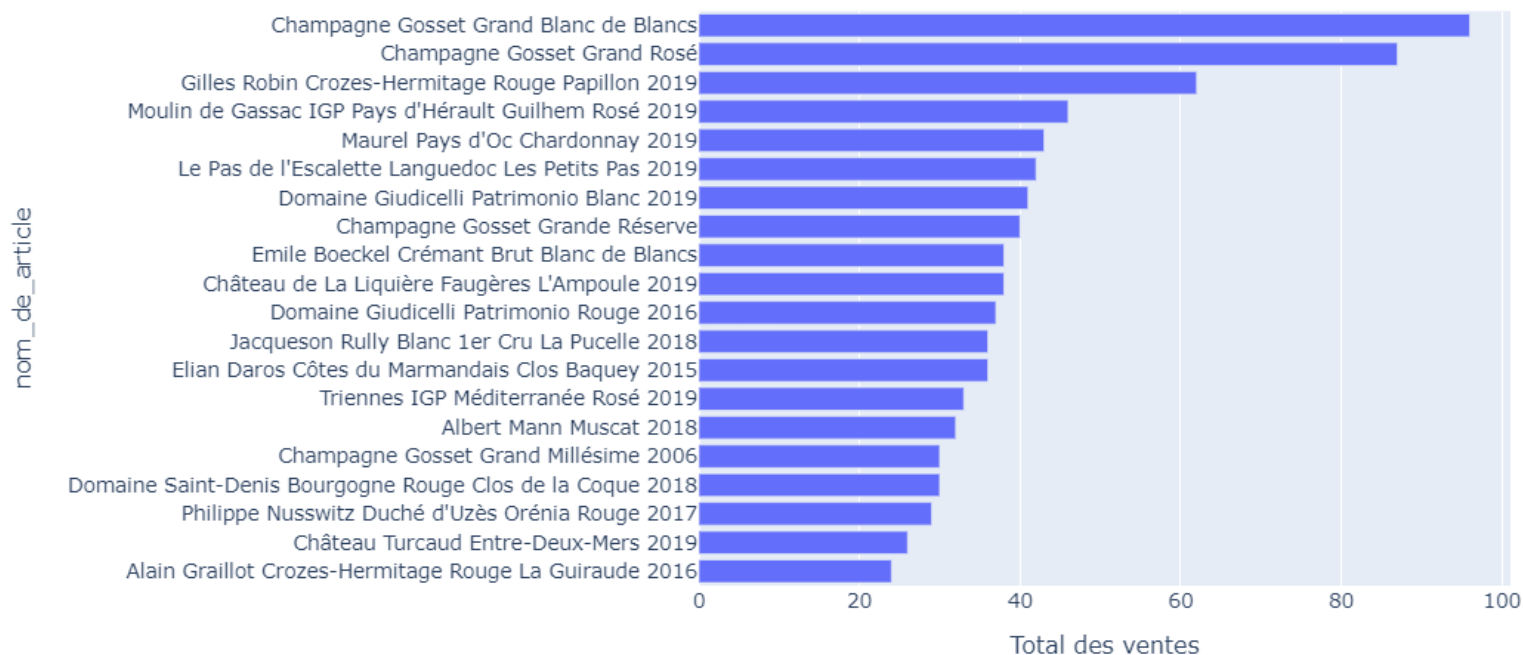
Palmares des 20 pemiers articles en CA



Analyse des ventes en Quantités

| Méthodes statistiques employés: | Graphique avec commentaire des résultats: | Limites éventuelles de l'analyse : |
|--|---|--|
| Clacul Total CA :70568.6€ | En termes de quantité, on peut clairement constater que le produit "Champagne Gosset Grand Blanc de Blancs" est celui qui se vend le plus, avec notamment 96 ventes totales | Certain produits pouvant potentiellement fausser l'analyse |
| Calcul du20/80 en CA: 351 articles=80% | | |

Classement des 20 premiers articles par quantité vendue



Actions pour la suite

- Procéder à la verification avec l'équipe technique des 'SKU' des articles
- De même, ajuster l'attribut 'id_web' des articles

Point sur les compétences apprises

- *Qu'est-ce qui s'est bien passé pour vous dans ce travail de nettoyage ?*
 - Le processus de nettoyage des données s'est bien déroulé grâce à l'utilisation
 - Python
 - La bibliothèque Pandas.
 - La bibliothèque Numpy
- *Qu'est-ce que vous avez trouvé le plus difficile ?*
 - La gestion des données manquantes
 - Des valeurs aberrantes
 - La cohérence des références produit entre l'ERP et la boutique en ligne
- *Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ?*
 - L'utilisation la bibliothèque Plotly

MERCI!