

PRESENTATION DU RAPPORT D'ANALYSE DE BOTTLENECK

PRÉSENTÉ PAR : MILAINE GUIAGAING

BUSINESS INTELLIGENCE ANALYST

DATE : 14/12/2023

ENJEUX ET OBJECTIFS

Consolider les données de l'ERP et du site de vente en ligne afin d'analyser les ventes.

OBJECTIF 1: FAIRE LE RAPPROCHEMENT ENTRE L'ERP ET L'EXPORT DE LA TABLE DE LA BOUTIQUE EN LIGNE;

OBJECTIF 2: ANALYSER LA VARIABLE PRIX AFIN DE DÉTECTER LES VALEURS ABERRANTES, LES LISTER ET LES REPRÉSENTER GRAPHIQUEMENT;

OBJECTIF 3: CALCULER LE CHIFFRE D'AFFAIRES PAR PRODUIT, AINSI QUE LE TOTAL DU CHIFFRE D'AFFAIRES RÉALISÉ EN LIGNE.

PLAN

1. ANALYSE EXPLORATOIRE DES FICHIERS
2. NETTOYAGE DES DONNÉES
3. CONSOLIDATION DES DONNÉES
4. ANALYSES UNIVARIÉES
5. ACTIONS POUR LA SUITE
6. COMPETENCES APPRISES

1. ANALYSE EXPLORATOIRE DES FICHIERS

Df_erp: références produit, leur prix de vente et leur état de stock

- Il n'y a aucune ligne à null, pas de doublons et la clé primaire est bien unique.
- Le nombre de lignes et de colonnes (825,5)
- Aucune difficulté.

Df_web: les informations des produits commercialisés en ligne

- La colonne 'sku' correspond à l'id_web du fichier liaison
- La clé primaire n'est pas unique (présence de valeurs nulles).
- Le nombre de lignes et de colonnes (714,5)

Df_liaison: liste les product_id de l'ERP avec leur référence côté Web

- La clé de référence id_web contient des valeurs nulles.
- Le nombre de lignes et de colonnes (825,2)

Df_caracteristiques: caractéristiques des alcools commercialisés

- Le nombre de lignes et de colonnes (611,13)

Suppression de la variable « stock_status » car redondante avec la variable stock_quantity

- Suppression des colonnes qui avaient peu d'informations
- Suppression des doublons (798) et valeurs nulles (85)
- Uniformiser les valeurs 'bon-cadeau-25-euros', '13127-1'.

2. NETTOYAGE DES DONNÉES (1/2)

- DÉTERMINER LA TYPOLOGIE DES DONNÉES
- VÉRIFIER L'UNICITÉ DE LA CLÉ PRIMAIRE
- DÉTECTER LES DOUBLONS ET LES SUPPRIMER

```
RangeIndex: 825 entries, 0 to 824  
Data columns (total 5 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   product_id      825 non-null    int64  
1   onsale_web      825 non-null    int64  
2   price           825 non-null    float64  
3   stock_quantity  825 non-null    int64  
4   stock_status    825 non-null    object  
dtypes: float64(1), int64(3), object(1)
```

```
doublons=df_liaison['product_id'].duplicated().sum()  
print(f"Le nombre de doublons est :{doublons}, donc les valeurs de cette colonne sont toutes uniques")  
  
Le nombre de doublons est :0, donc les valeurs de cette colonne sont toutes uniques
```

```
doublons =df_web['id_web'].duplicated().sum()  
print(f"Le nombre de doublons est : {doublons}")  
  
Le nombre de doublons est : 798  
df_web.drop_duplicates(inplace=True)
```

2. NETTOYAGE DES DONNÉES (2/2)

- DÉTERMINER LA PERTINENCE DE CHAQUE VARIABLE
- RECHERCHER ET CORRIGER LES VALEURS NE RESPECTANT PAS LA CODIFICATION DE LA VARIABLE
- SUPPRIMER LES VARIABLES QUI N'AVAIENT PAS D'INTÉRÊT POUR NOTRE ANALYSE

Le nombre de valeurs présentes dans chaque colonne est :

id_web	714
virtual	1
downloadable	1
rating_count	1
average_rating	1
total_sales	41
tax_status	1
tax_class	0
post_author	2
post_date	716
post_date_gmt	716
post_content	0
post_title	713
post_excerpt	679
post_status	1
comment_status	1
ping_status	1
post_password	0

```
for idx,val in df_web['id_web'].dropna().items():  
    if type(val)!= int:  
        print(idx,val)
```

```
0 bon-cadeau-25-euros  
797 13127-1  
1209 bon-cadeau-25-euros  
1511 13127-1
```

```
df_erp.drop(columns=['stock_status','stock_status_2'],inplace=True)
```


3. CONSOLIDATION DES DONNÉES

df_erp et df_liaison

- Jointure interne sur la clé primaire « product_id ».

df_merge et df_web

- Jointure interne sur la clé primaire « id_web » pour éviter des Nan car le fichier merge contenait des lignes sans correspondances dans le fichier Web (825 contre 714).

df_merge et df_caractéristiques

- Jointure à gauche (left_join) sur la clé primaire « post_name » afin de conserver toutes les lignes du df_merge.

4. ANALYSE UNIVARIÉE (1/3)

□ Du prix

La moyenne du prix est : 32.49

L'ecart-type du prix est : 27.81

Nombre outliers : 32

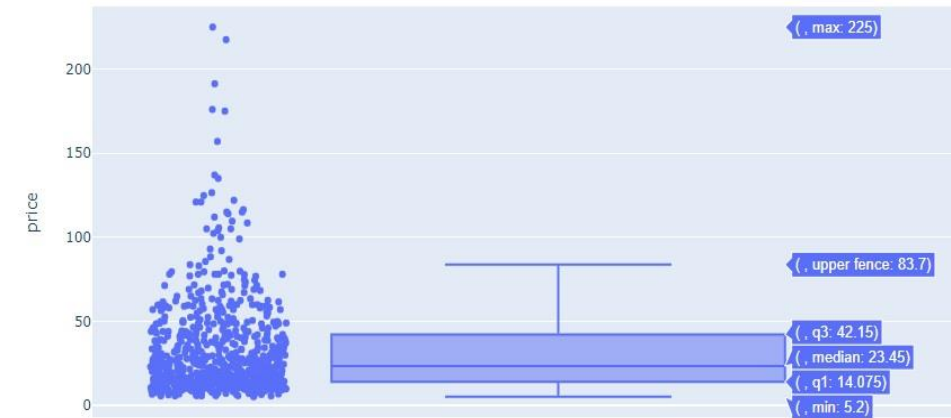
Proportion d'articles outliers : 4.20

Seuil de prix pour les Outliers: 85.6

Aucune valeur manquante ou d'article sans prix renseigné.

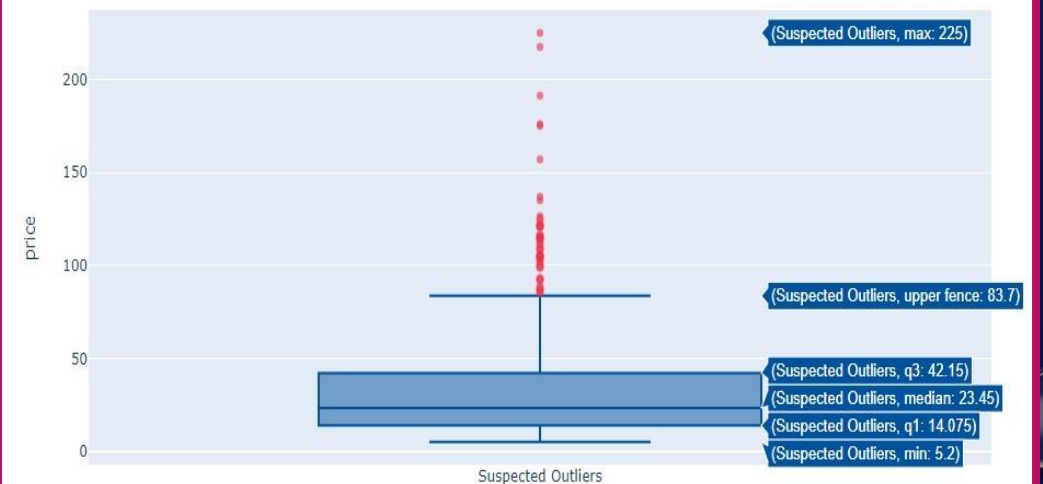
Bête à moustache-Repartition des prix

Boîte à moustache de la repartition des prix



Bête à moustache-Repartition des Outliers

Boxplot des Outliers



4. ANALYSE UNIVARIÉE (2/3)

□ Du chiffre d'affaires

Le chiffre d'affaires total : 70568.6 euros

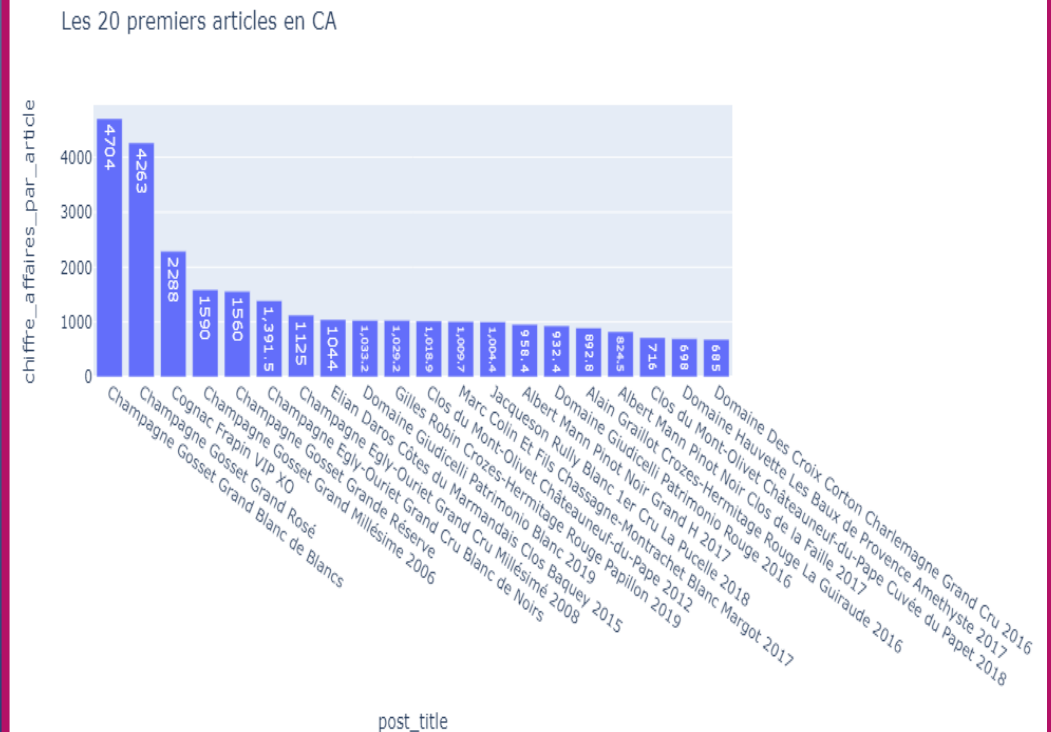
Les 20 articles avec les CA plus élevés sont :

Le nombre d'articles représentant 80% du CA est : 130

Leur proportion est: 18.20%

La loi de Pareto est bien respectée.

Les 20 premiers articles en CA



4. ANALYSE UNIVARIÉE (3/3)

□ DES VENTES

Les 20 articles avec les nombres de vente plus élevés sont :

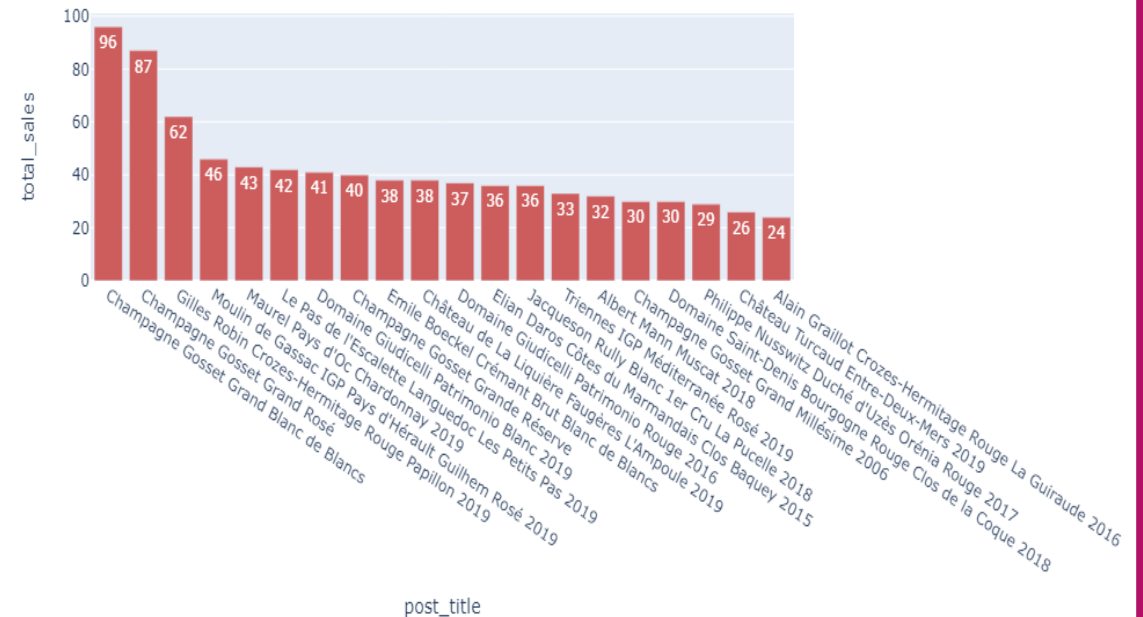
Le nombre d'articles représentant 80% du CA est : 150

Leur proportion est : 21.0%

La loi de Pareto est bien respectée.

Les 20 premiers articles en vente

Les 20 premiers articles en vente



5. ACTIONS POUR LA SUITE

- Faire une analyse bivariée, corrélation entre le nombre de vente et le chiffre d'affaire;
- Calculer le chiffre d'affaire par catégorie de produits, par région et par domaine;
- Veiller à la conformité de la base de données;
- Se rapprocher des équipes afin d'en savoir davantage sur les valeurs aberrantes observées et les informations manquantes.

6. COMPÉTENCES APPRISSES

- Nettoyer une base de données avec python;
 - Réaliser une analyse exploratoire des données;
 - Consolider des données afin d'en tirer des informations pertinentes;
 - Réaliser une analyse univariée.
-
- *J'ai eu besoin de plus d'entrainement pour le calcul du 20/80*