

RAPPORT D'ANALYSE: BOTTLENECK

Amina Feredj - Consultant BI
Paris, le 20 décembre 2023

Les étapes

Importation des librairies nécessaires

Importation et création des fichiers

Analyse exploratoire: Principaux codes et résultats

Nettoyage et consolidation des données

Analyses univariées et chiffres clés

Problématiques rencontrées et axes d'amélioration

Compétences acquises

IMPORTATION DES LIBRAIRIES

- L'importation des librairies se fait via la commande « import » suivi du nom de la librairie.
- On peut ajouter « as » pour définir un alias.

Pandas:

Pour manipuler et analyser des données

Numpy:

Pour faire des opérations logiques ou mathématiques sur des tableaux

Chardet:

Pour identifier l'encodage du fichier

Plotly express:

Pour crée des visualisations de données interactives

Statscs:

Pour calculer des valeurs statistiques avec des données numériques

IMPORTATION ET CREATION DES FICHIERS ⁴

- L'importation d'un fichier via la commande « `pd.read` » et en fonction de son extension (Ex: `df_erp = pd.read_excel (« erp.xlsx »)`)
- Le codes (`pd.options.display.max_columns = None`) et (`pd.options.display.max_rows = None`) permettent d'afficher toutes les lignes et colonnes

df_erp:

Références des articles, prix et quantités

df_web:

Informations sur les articles vendus en ligne

df_caracteristiques:

Caractéristiques des articles commercialisés

df_liaison:

Rapprochement des identifiants entre `df_erp` et `df_web`

df_merge:

Fusion de tous les fichiers

ANALYSES EXPLORATOIRES⁵

PRINCIPAUX CODES

Nombre
d'observations

```
format(dataset.  
shape[0])
```

Nombre de
colonnes

```
dataset.columns
```

Nature des
données

```
dataset.dtypes
```

Valeurs par
colonne

```
dataset.info()
```

Nombre de
doublons

```
dataset['colonne'].  
duplicated().sum()
```

La valeur
minimum

```
dataset['colonne']  
.min()
```

La valeur
maximum

```
dataset['colonne'].  
max()
```

Afficher les
premières lignes

```
Dataset.head()
```

Valeurs uniques

```
dataset['colonne'].  
unique()
```

Fichier df_erp

Nombre d'observations: 825

Nombre de colonnes: 5

Nature des données: int64, float64 et object

Nombre de doublons et NAN : 0

- La clé primaire est unique
- Les colonnes stock_status et stock_quantity sont redondantes

Nombre d'observations: 1513

Nombre de colonnes: 28

Nature des données: object, int64, float64, datetime64

Nombre de doublons: 714

- Les valeurs des colonnes « sku » et « id_web » correspondent
- La clé primaire n'est pas unique (Valeurs nulles 85 et codification différente 4)

Fichier
df_web

Fichier df_liaison

Nombre d'observations: 825

Nombre de colonnes: 2

Nature des données: int64, object

Nombre de doublons: 0

-
- Présence de 91 valeurs nulles (id_web)
 - Valeurs uniques (product_id)
-

Nombre d'observations: 611

Nombre de colonnes: 13

Nature des données: object , float64

-
- Présence de valeurs nulles
 - Présence de plusieurs colonnes d'informations
-

Fichier
df_caracteristiques

NETTOYAGE ET CONSOLIDATION DES DONNÉES

Identifier la nature des valeurs et leur pertinence

Vérifier l'unicité de la clé primaire et supprimer les doublons

Identifier les valeurs nulles et les remplacer

Harmoniser la codification (ex: id_web : bon-cadeau-25-euros et 13127-1)

Supprimer les colonnes redondantes (ex: stock_status et stock_quantity)

Renommer la colonne « sku » en « id_web » pour faciliter la consolidation

NETTOYAGE ET CONSOLIDATION DES DONNÉES

Consolidation des fichiers en 3 étapes :

- **df_erp et df_liaison** : Jointure interne car la colonne « product_id » est commune aux deux fichiers
- **df_merge et df_web** : Jointure interne car la colonne « id_web » clé primaire est commune aux deux fichiers.

Eviter d'avoir des valeurs nulles, le fichier df_merge contient des lignes sans correspondances dans le fichier df_web.

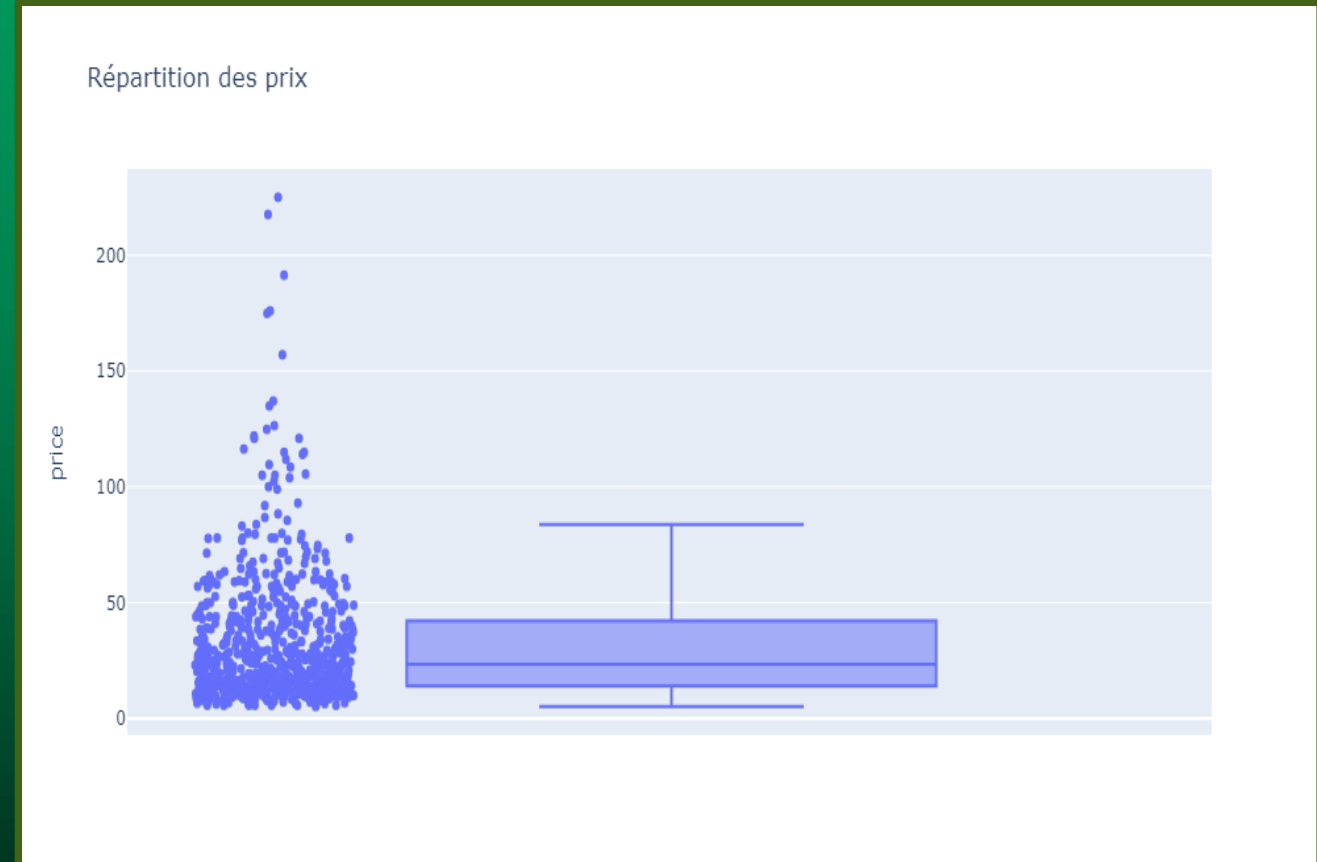
- **df_merge et df_caractéristiques** : Jointure à gauche car la clé primaire « post_name » est commune aux deux fichiers.

Trouver les correspondances aux lignes du fichier df_merge.

ANALYSE UNIVARIEE ET CHIFFRES CLÉS

- **LE PRIX:**

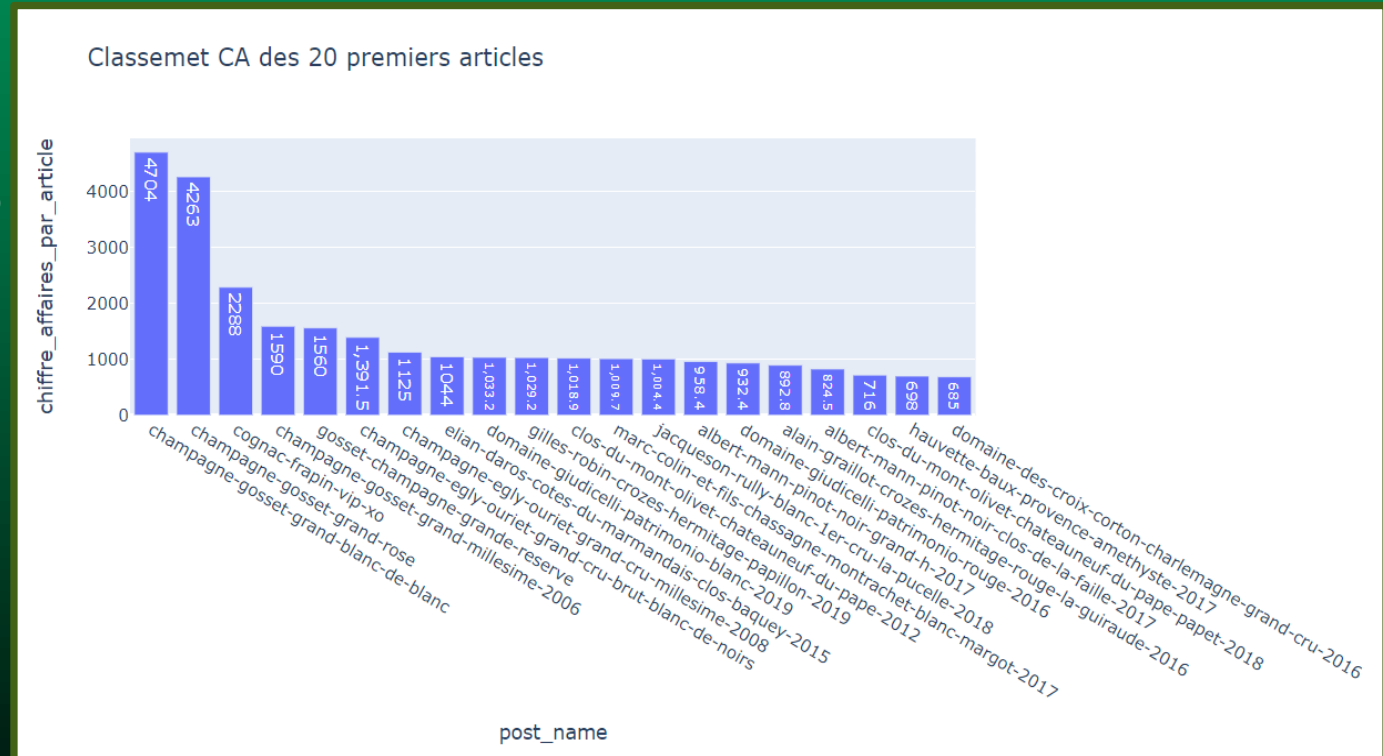
- Moyenne du prix : 32.49
- Ecart-type : 27.8
- Nombre de « Outliers » : 32
- Proportion d'articles « Outliers » : 4.48
- Seuil de prix pour les « Outliers »: 84.35



ANALYSE UNIVARIEE ET CHIFFRES CLÉS

• LE CHIFFRE D'AFFAIRES:

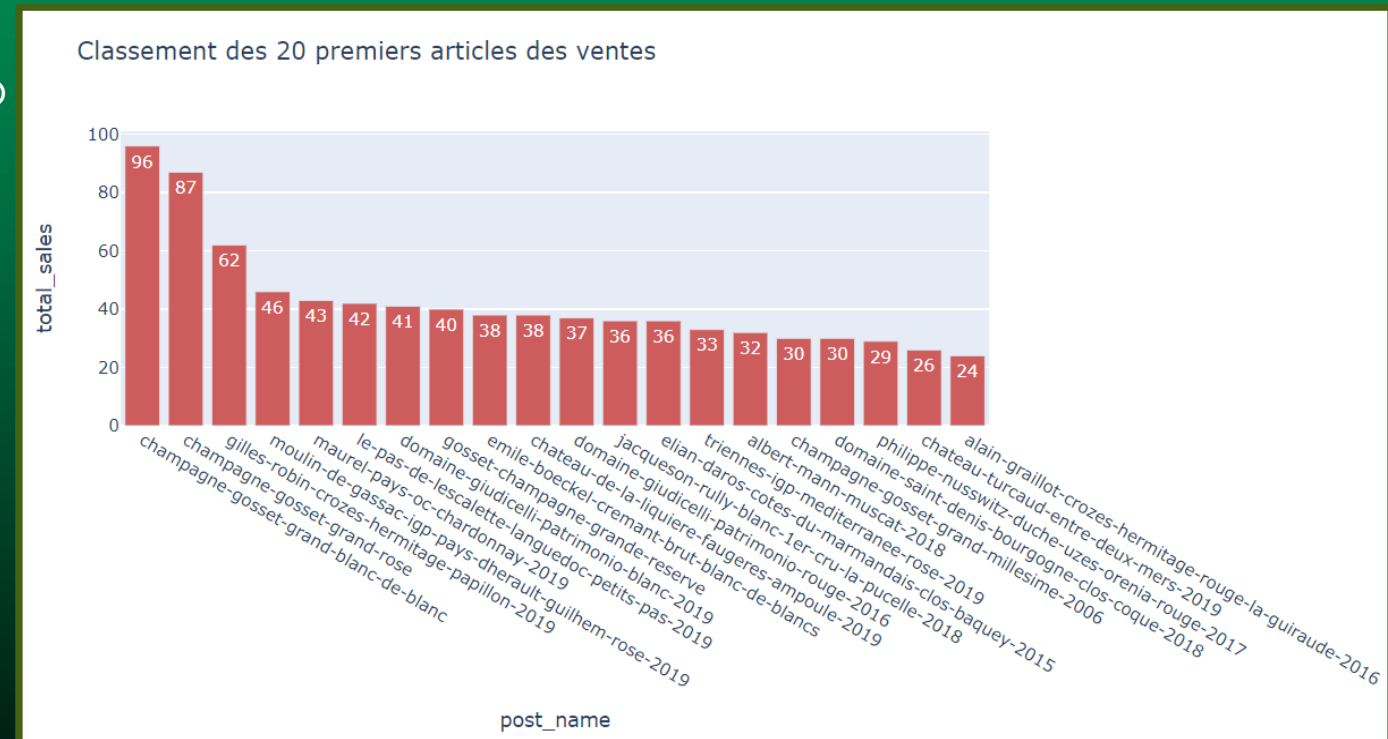
- Chiffre d'affaires (CA)total : 70568.6
- Nombre d'articles représentant 80% du CA : 163
- Proportion dans le catalogue : 23%
- Les chiffres illustrent bien la loi de Pareto



ANALYSE UNIVARIEE ET CHIFFRES CLÉS

• LE NOMBRE DE VENTES:

- Nombre d'articles représentant 80% des ventes : 150
- Proportion dans le catalogue : 21%
- Les chiffres illustrent bien la loi de Pareto



PROBLÉMATIQUES RENCONTRÉES ET AXES D'AMÉLIORATION

Problématiques rencontrées:

- Obtention de plusieurs messages d'erreurs
- Identification des bonnes variables pour les calculs
- Difficulté à trouver les bons codes

Axes d'amélioration:

- Essayer de cerner les « Outliers » et informations manquantes
- Uniformiser les codifications au préalable
- Faire des analyses bivariées pour trouver des corrélations pertinentes (Exemple: entre le panier moyen et le chiffre d'affaires)
- Faire des agrégations par famille ou région par exemple.
- Optimiser le temps et les efforts

COMPÉTENCES ACQUISES

- ✓ Identifier le bon type de jointure pour éviter les doublons
- ✓ Vérifier les résultats suite à l'exécution des codes
- ✓ Nettoyer une base de données avec python
- ✓ Trouver les chiffres clés grâce à des analyses exploratoires pertinentes
- ✓ Consolider des données réparties sur plusieurs fichiers
- ✓ Faire des analyses univariées.
- ✓ Identifier les messages d'erreur et les corriger