

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

Léo Morel

[Business analyst]

[Date de la présentation]

Analyses Exploratoires des Données

Informations données :

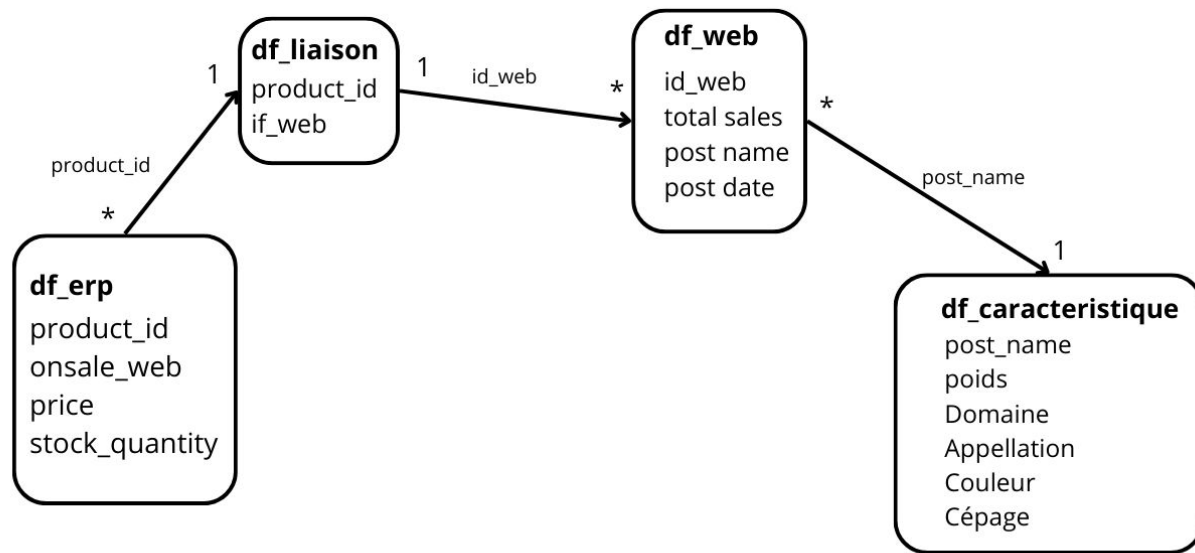
- **Datasets :**
 - 4 datasets (Erp, Web, liaison, caractéristique)
- **Caractéristiques**
 - Plusieurs éléments vides
 - beaucoup de champs inutiles

Remarques, difficultés rencontrées :

- *Nettoyages des données*
 - *Suppression des lignes vides, des valeur incohérentes*
 - *identification des “Outliers”*
- *Features engineering*
 - *Utilisation de moyenne*
 - *Création de colonne (somme cumulative, proportion)*

Schéma de la base de donnée

Shéma Base de données après traitement



Traitement réalisés

Nettoyages des données

- **Suppression des lignes vides**
 - Vérification des lignes ayant des informations manquantes
 - Suppression des lignes avec trop peu d'éléments
- **Identification des “Outliers”**
 - Utilisation du :
 - Z-score
 - Upper Fence
 - par déduction
 - vérifiez si les “Outliers” sont coherent

Features engineering

- **Agrégation de données par groupe**
 - Création de dataframe par groupe (Région, Domaine, couleur)
- **Création de colonne :**
 - “ca_par_article”, le chiffre d'affaire par article
 - “part_ca”, calcul la part du CA par article
 - “somme_cumulative”, addition cumulative de la dernière colonne

Fusion ou consolidations des données

Choix éléments analyse :

- *Choix des attributs*
 - Variable “price”
 - Variable “CA”
- *Clés utilisés*
 - “Product_id”
 - “id web”
 - “post_name”

Difficultés rencontrées :

- *Vigilances particulières au cours du traitements*
 - Nombre de lignes obtenus
 - Duplications de données
 - valeurs vides
- *Difficultés ou pièges rencontrés*
 - Effectué tous les transformations des fichiers avant fusion
 - Aller trop vite

Analyses univariées du prix

Méthodes statistiques employés

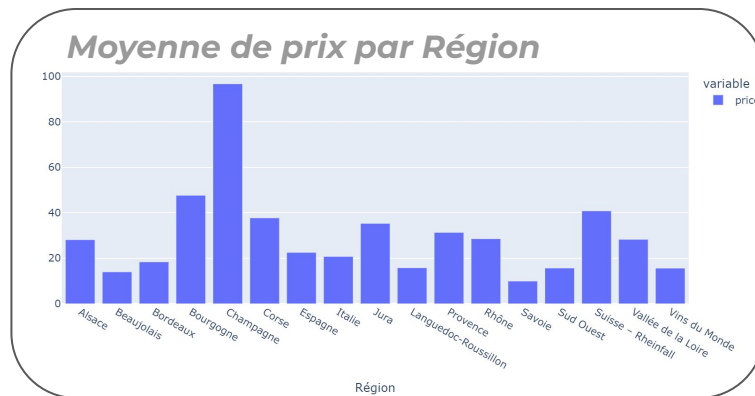
- utilisation de moyenne, médiane, fréquence

Graphique des résultats

- La moyenne des prix est **28 €**
- le prix le plus fréquent est le **11€**
- la médiane des prix est **~21 €**

Commentaire des résultats

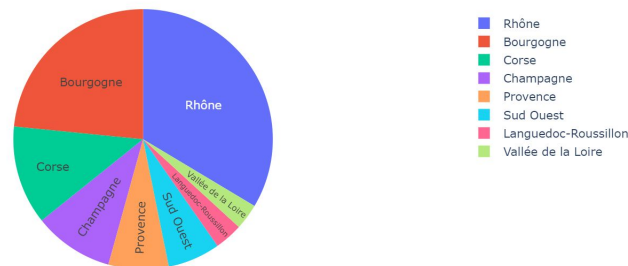
- Le prix de la majorité des produits vendus, se situe aux alentours de **11 €**
- Le prix maximum est **217,5 €**, et le minimum **5,2 €**



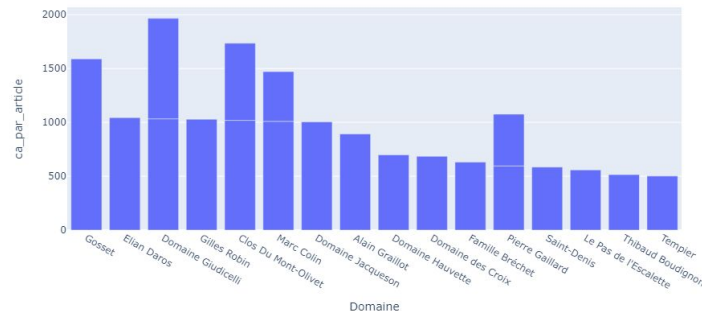
Analyses univariées du CA

- **Méthodes statistiques employés**
 - Triage du dataset par articles avec les meilleurs CA
 - Agrégation des données avec différents champs (domaine, région, article)
- **Graphique avec commentaire des résultats**
 - 4% des produits font 80 % du Chiffre d'affaire
 - Le CA du site est de **41 804 €**
 - La région qui fait le plus de CA est le **Rhône**
 - La région qui fait le moins de CA est **la Vallée de la Loire**
 - Le domaine qui a le plus de CA est **"Domaine Giudicelli"**

Chiffre d'affaire par Région



Chiffre d'affaire par Domaine



Actions pour la suite

Mettre en place des dataset avec :

- *moins d'information inutiles (garder que les informations nécessaires)*
- *éviter les erreur de saisies*

Mettre en place des rapports Power BI pour mieux présenter les résultats :

- *Présenter le chiffre d'affaire par région sur une carte*
- *Etablir des objectifs de ventes pour mieux situer l'activité*

Créer une Base de donnée avec SQL :

- *Permet une meilleur sécurité de la base de données*
- *Une meilleur logistique de stockage des données pour une meilleur exploitation*

Point sur les compétences apprises

Qu'est-ce qui s'est bien passé pour vous dans ce travail de nettoyage ?

- le nettoyage de données (valeurs vides, duplicatas)

Qu'est-ce que vous avez trouvé le plus difficile ?

- vérifiez que les résultats obtenus sont fiables et cohérents
- l'organisation du notebook (on peut se perdre facilement)

Sur quelles tâches est-ce que vous pensez avoir besoin de plus d'entraînement ?

- sur l'analyse des résultats (commenter les résultats)
- présenter ces résultats avec des graphiques plotly
- Trouver les indicateurs d'analyse les plus pertinents