

Rapport d'analyse business

Ce rapport présente les différents traitements et analyses réalisés dans mon notebook en tant que business analyst. Il couvre la mise en place de l'environnement, l'analyse exploratoire des données, la fusion des fichiers, l'analyse de cohérence et la data visualisation.

F

par Flavien Bardet

Mise en place de l'environnement et préparation des données



Importation des librairies

Import panda, plotly, seaborn, numpy, matplotlib

Configuration de l'affichage

```
pd.set_option('display.max_columns', None)
```

Chargement des fichiers

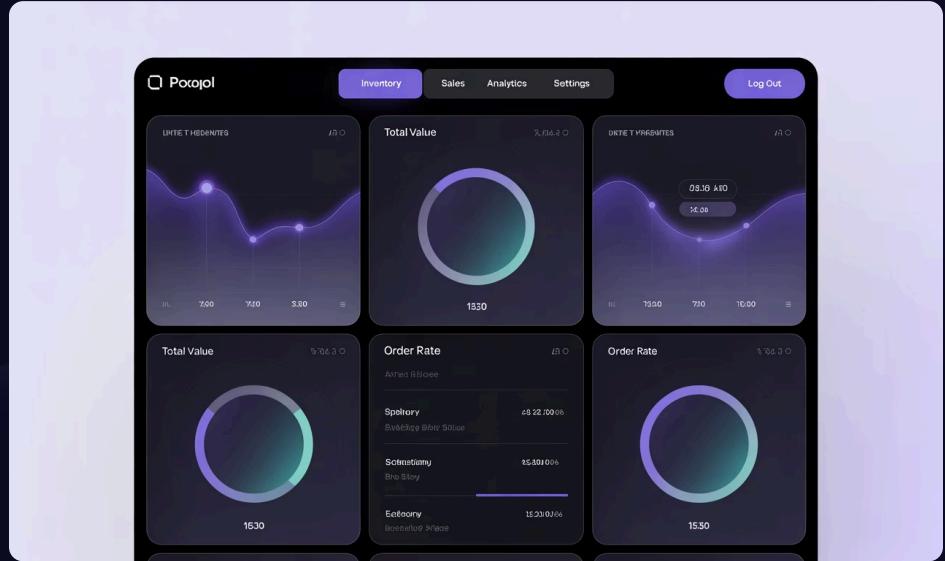
Chargement des fichiers web, erp, liaison

La première étape de notre analyse consiste à mettre en place l'environnement de travail en important les bibliothèques nécessaires, en configurant l'affichage pour visualiser toutes les colonnes des dataframes, et en chargeant les trois fichiers de données qui seront utilisés tout au long de l'analyse.

Analyse exploratoire du fichier ERP

Caractéristiques générales

- Dimensions: 6 colonnes, 825 observations
- Pas de doublons dans product_id
- Valeurs de stock_status: instock, outofstock
- Stock_status varie avec stock_quantity



Anomalies détectées et corrections

| Type d'anomalie | Détails | Action corrective |
|---------------------------|--|--|
| Incohérences stock_status | ID: 4039 stock 3 signalé outofstock ID: 4885 stock 0 marqué instock | Alignement stock_status sur stock_quantity |
| Prix négatifs | 3 produits: 4233 à -20, 5017 à -8, 6594 à -9,1 | Suppression des prix négatifs |
| Stock négatif | 2 anomalies: 4973 à -10, 5700 à -1 | Suppression des stocks négatifs |

L'analyse du fichier ERP a révélé plusieurs anomalies qui ont été corrigées pour assurer la qualité des données. Une colonne stock_status_2 a été créée temporairement pour vérifier la cohérence entre le statut de stock et la quantité disponible, puis supprimée après correction des incohérences.

Concernant les prix d'achat, aucune anomalie n'a été détectée avec des valeurs allant de 2,74€ à 137,81€. La répartition des valeurs onsale_web montre que 716 produits sont disponibles à la vente sur le site (valeur 1) contre 109 qui ne le sont pas (valeur 0).

Analyse exploratoire du fichier Web



Dimensions

29 colonnes et 1513 observations



Sélection des colonnes

Conservation uniquement des colonnes pertinentes pour l'analyse



Anomalies SKU

85 valeurs Nan et 2 anomalies de format: 13127-1, bon-cadeau-25-euros

Analyse des SKU

Les règles de codification du SKU ont été définies avec un format attendu de chiffres entre 1 et 20 caractères (`regex_sku_valide = r"^\d{1,20}$"`). L'analyse a révélé:

- 85 valeurs Nan dans les SKU
- 2 anomalies de format: 13127-1, bon-cadeau-25-euros
- 1513 SKU dupliqués
- 714 SKU distincts qui apparaissent chacun 2 fois

L'analyse du contenu des SKU non renseignés a montré que 83 observations sont entièrement vides (anomalies à supprimer) et 2 contiennent des données partielles nécessitant une clarification métier.

Analyse du fichier liaison et fusion des données

Caractéristiques du fichier liaison

- Dimensions: 2 colonnes, 825 observations
- id_web: 734 valeurs non nulles (type object)
- product_id: 825 valeurs non nulles (type int64)
- 91 valeurs manquantes dans id_web

Processus de fusion

1. Fusion left join entre df_erp et df_liaison: df_merge
2. Nettoyage des lignes vides dans df_web (85 lignes avec SKU = Nan)
3. Dédoublement de df_web avec post_type = product
4. Filtrage de df_merge pour exclure les liens sans id_web valide
5. Joindre inner join entre df_merge et df_web: df_final
6. Suppression des colonnes inutiles (id_web et colonnes temporaires)

Caractéristiques des dataframes après traitement

| Dataframe | Colonnes | Observations |
|------------------------|----------|--------------|
| df_web post traitement | 7 | 715 |
| df_merge post filtrage | 8 | 734 |
| df_final | 15 | 714 |

Analyse de la cohérence des données



Cohérence entre erp et liaison

Vérification des web_id manquants et produits offline



Cohérence entre erp et final

Vérification des produits offline dans le fichier web



Cohérence entre erp et web

Recherche de produits web only

Anomalies identifiées

- 21 produits supposés offline mais avec id_web
- 3 produits marqués online mais sans id_web: 4594, 5070, 5075 (explication probable: produits sans stock)
- 1 produit marqué offline présent dans le fichier web: Product_id: 4200, sku: 16295 (CA: 81,2€)

Comparaison nombre d'items

La table erp contient 716 produits marqués online, tandis que la table final en contient 714. Cette différence s'explique par:

- 3 produits marqués online absents de final: 4594, 5070, 5075
- 1 produit marqué offline présent dans final: 4200

L'analyse n'a pas révélé de produits existant uniquement dans le fichier web (web only).

Analyse des prix et du chiffre d'affaires

Statistiques des prix

- Moyenne: 32,33€
- Écart type: 27,60€
- Z-score 3: 115,12€
- Nombre d'outliers: 31 articles (4,34% du catalogue)

L'analyse des outliers montre que pour les prix élevés, il s'agit de produits de luxe donc les prix sont cohérents.

Les prix négatifs ont été corrigés comme anomalies.



Analyse du chiffre d'affaires

Le chiffre d'affaire total du site web s'élève à 143.680,10€.

143,680€

CA Total

Chiffre d'affaires global du site web

435

Produits clés

Nombre d'articles représentant 80%
du CA

60.92%

Proportion

Pourcentage du catalogue
représentant 80% du CA

L'analyse de Pareto montre que 435 articles (60,92% du catalogue) génèrent environ 80% du chiffre d'affaires total, ce qui indique une concentration relativement importante des ventes sur une partie du catalogue.

Analyse des stocks et des marges

Analyse des stocks

- Temps de rotation global des stocks: 3,44 mois
- Quantité totale de stock disponible en ligne: 16740
- Valorisation totale des stocks: 277.328,07€
- 3 produits dormants identifiés: 4337, 4355, 5932 (2 champagnes, un cognac)

Analyse des marges

- Taux de marge min: -634,99% (anomalie)
- Taux de marge max: 47,76%
- Produit avec taux de marge négatif: ID 4355, prix achat 77,48€, prix vente (TTC) 12,65€
- Nombre de produits avec marge positive: 713
- Marge positive min: 22,78%
- Marge positive max: 47,76%

Corrélations entre stock, prix et ventes

Une analyse des corrélations a été réalisée pour identifier les relations entre les niveaux de stock, les prix et les ventes. Cette analyse a été visualisée à l'aide d'une heatmap de corrélation.

L'analyse a également révélé que nous vendons de l'huile d'olive, mais que ce produit n'est pas très rentable en termes de marge.

La visualisation du top 20 des produits par quantité vendue et l'analyse Pareto montrent que 434 articles (60,78% du catalogue) représentent environ 80% des ventes en quantité, ce qui est cohérent avec l'analyse du chiffre d'affaires.