

ESG Finance

Machine Learning

TP2

Online Retail Transactions Analysis
`#data-analysis, #machine-learning`

Novembre 2022

1 Objectifs

- Utiliser Python et les librairies usuelles¹ pour analyser, visualiser², résumer des informations dans de la donnée transactionnelle
- Utiliser Python et les librairies usuelles pour prédire des ventes

2 Données

Vous disposez de 25 900 dépenses. Les données sont stockées sous la forme d'un fichier Excel à l'adresse suivante : https://docs.google.com/spreadsheets/d/1sdiB9aZkP1k9deEFfT-HwCGwtCSP3pd-/edit?usp=share_link&ouid=106425905375102751639&rtpof=true&sd=true

Le fichier contient 8 colonnes :

- InvoiceNo : identifiant de facture associée à une dépense ou panier d'achat
- StockCode : identifiant de produit
- Description : description du produit au format textuel
- Quantity : quantité de produits achetés
- InvoiceDate : date de l'achat

1. <https://pypi.org/project/sqldf/>

2. <https://matplotlib.org/>, <https://seaborn.pydata.org/>

- UnitPrice : prix unitaire du produit
- CustomerId : identifiant de l'acheteur
- Country : pays où l'achat a été réalisé

3 Questions (10 points)

- **Volume.** Lors d'une dépense, combien de produits un client achète-t-il généralement ? Tracez la distribution associée. Faites apparaître le nombre moyen et le nombre médian de produits acheté lors d'une dépense, par pays.
- **Montant.** Lors d'une dépense, quel est le montant dépensé par un client généralement ? Tracez la distribution associée. Faites apparaître le montant moyen et le montant médian d'un panier client, par pays.
- **Volume.** Quels sont les top-5 produits les plus vendus durant le premier trimestre 2011 ? Pendant le second trimestre 2011 ?
- **Montant.** Quels sont les 5 pays qui ont générés le plus gros chiffre d'affaire pendant le premier trimestre 2011 ?
- **Montant.** Tracez l'évolution mensuelle du montant moyen et du montant médian d'un panier client pour chacun de ces cinq pays.
- **Fréquence.** A quelle fréquence dépensent les 100 plus gros clients du Royaume-Uni ?

4 Modèles (10 points)

- Quel sera le chiffre d'affaire global de la société en Décembre 2011 ?
- Combien de clients vont réaliser au moins un achat en Décembre 2011 en France ?

Pour conforter votre réponse, vous indiquerez la performance de votre modèle sur une période de test antérieur, e.g., Novembre 2011.