**Analyse des ventes**

**“RESTER LIVRE”**

**Contexte et description**

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, logo

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, document

Description générée automatiquement

Recommandations :

* Utiliser python et jupyter notebook
* Utiliser les librairies python de manipulation de data
* Bien structuré le code python dans le notebook (un box par action précise) et mettre sur le repo github au fur et a mesure
* Commenter vos résultats (de nettoyage et d’analyse) au fil de l’évolution

**Partie 1 : nettoyage des données**

* Traitement des valeurs manquantes
* Traitement des valeurs aberrantes et atypiques
* Analyse et correction éventuelle de chaque dataframes
* Traitement des éventuels doublons
* Conclure

**Partie 2 : les analyses**

* Décrire les variables des 3 dataframes (qualitatives et quantitatives)
* Analyse de l'évolution du chiffre d'affaires
* Analyse des prix pratiqués : variable 'price' (df\_transactionsv1\_products)
  + Mesures de tendance centrale et dispersion des prix produits (toutes catégories confondues)
  + Mesure de concentration des prix produits (toutes catégories confondues)
* Analyse sectorielle de l'activité : variable 'categ' (df\_transactionsv1\_products)
* Analyse des produits disponibles par catégorie
* Analyse démographique de l'activité
* Mesure de concentration des âges clients
* Analyse comportementale des clients
  + Analyse des transactions par client
  + Analyse des centres d'intérêt client

**Partie 3 : les corrélations**

**Les corrélations indiquent si deux variables sont linéairement équivalentes.**

Voici quelques questions supplémentaires pour nous aider à mieux comprendre les ventes enregistrées.

* Y a-t-il une corrélation entre le sexe des clients et les catégories de produits achetés ?
* Y a-t-il une corrélation entre l'âge des clients et le montant total des achats ?
* Y a-t-il une corrélation entre l'âge des clients et la fréquence d’achat (nombre d'achats par mois) ?
* Y a-t-il une corrélation entre l'âge des clients et la taille du panier moyen (en nombre d’articles) ?
* Y a-t-il une corrélation entre l'âge des clients et les catégories de produits achetés ?

**Un peu de théorie :**

Pour reprendre la définition de *Wikipédia* que je trouve plutôt bien trouvée :

"En probabilités et en statistique, la corrélation entre plusieurs variables aléatoires ou statistiques est une notion de liaison qui contredit leur indépendance."

* Préparation des données : Jointure des 3 dataframes
* Première approche des corrélations sur l'ensemble de nos variables numériques
* Analyse de la corrélation entre le sexe des clients et les catégories de produits achetés
* Analyse de la corrélation entre l'âge clients et le montant total des achats
* Analyse de la corrélation entre l'âge clients et la fréquence d'achat
* Analyse de la corrélation entre l'âge clients et la taille du panier moyen
* Analyse de la corrélation entre l'âge clients et la catégorie produits