Notes Projet 4

Question :

Pour transactions il y a :

* Des doublons à supprimer
* Des valeurs tests à supprimer (identifiées par id\_prod = ’T\_0’ ou date = ‘^test\_.+$’)
* Des dates d’une précision allant jusqu’à la fraction de minute

Il faut donc :

* Supprimer les doublons
* Supprimer les valeurs tests
* Ne conserver que les dates au format année-mois-jour

Pour le moment, j’ai un script pour les dates qui fonctionne. Les suppressions de doublons et de valeurs tests sont faites en amont.

Faut-il faire un script pour toutes les opérations ?

Autres fichiers :

* Customers Ok
* Products : une valeur négative supprimée donc Ok

Question : pour tous les fichiers faut-il identifier les erreurs à nettoyer et créer un script même s’il n’y a pas d’erreurs?

Analyse :

Problèmes rencontrés :

1. Dans chapitre ‘Création du tableau ‘data’ pour l’analyse’ :
   * Colonne ‘Unnamed :0’ lors de l’import du csv. Tentative de résolution en rajoutant le paramètre index=False lors de l’export. Ne fonctionne pas avec ‘customers’.
   * Comment éviter l’apparition de ces colonnes ?
2. Dans sous-chapitre ‘Chiffre d’affaire mensuel’ :
   * Dates : Les mois sont numérotés de 01 à 12 dans a colonne ‘month’. La période étudiée va de mars 2021 à février 2022. Si je crée un graphique avec les mois en abscisse, les mois de janvier et février 2022 vont apparaître avant les mois de 2021. J’ai remplacé les valeurs de mois 1 et 2 par 13 et 14 pour respecter. Y a-t-il une autre méthode ?
3. Dans sous-chapitre ‘Chiffre d’affaire par catégorie et par mois’ :
   * Histogramme : comment afficher les colonnes de chaque catégorie empilées ?
4. Dans sous-chapitres ‘Top 10 mensuel des meilleures ventes et meilleurs clients’ :
   * Est-ce possible de créer une fonction ? (je fais une tentative dans le sous-chapitre ‘Top 10 des meilleurs clients par mois’)
5. Chaque fois que je crée un graphique (premier graphique dans sous-chapitre ‘Chiffre d’affaire mensuel’) :
   * Un message d’erreur apparaît même si l’affichage du graphique fonctionne.
   * Est-ce normal ?
6. Dans sous-chapitre ‘Courbe de Lorenz et indice de Gini’ :
   * L’export de la courbe de Lorenz donne un graphique sans courbe.
   * Quel est le paramètre à renseigner pour que l’export se fasse correctement ?

À faire :

* Faire export transparents ??
* Comparer graphiques actuels avec ceux fournis pour projet 2
* Faire un graph en camembert pour CA par catégorie => Export trop long
* Exporter Courbe de Lorenz
* Trouver source d’interprétation Indice de Gini
* Est-ce qu’il faut laisser courbe de distribution?