Travail effectué : new data\_set à partir des données de Promos et Reabos concaténés.

Une image contenant texte, nombre, Police, reçu

Description générée automatiquement

1ère étape : Reformatage des données

* Suppression des données Promos où il y avait deux promos qui se chevauchait à l’aide de la fonction drop\_dupplicated\_columns\_df\_Données\_Promos

Les promos ont été enlevés par ordre de priorités

Par exemple : si une même promo pour SG et ODD 15 jours TC on garde ODD 15 jours TC

* Join table Promos et Reabos : enft il y avait un gros problème dans notre ancien Join : on ne prenait pas en compte les personnes qui se réabonnait plusieurs fois dans le Join ID\_ABONNE / DATE\_ACTE\_REEL le même jour à des promos (par exemple bénéficie d’une promo qui commence le 15 juin 2021 et une autre le 30 juin 2021 mais la date d’acte réel le 15 juin 2021) différentes et cela donnait un truc comme ça :

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Ces personnes représentaient à peu près 3% des abos plus de deux fois le même jour et seulement 0,005 % plus de trois fois le même jour. Est-ce que cela est significatif ou juste les drop ?)

Tentative de palier à cela un peu infructueuse :

1. Prendre en compte : DATE\_PRISE\_EFFET (pour le REABO) = DATE\_DEMARAGE\_PROMO (pour la PROMO pour le join cela marchait bien à part si la promo commençait avant et on avait perdu beaucoup de Semaines Généreuses

Ensuite, Rajouter ceux qui bénéficiais d’une promo avant le début du REABO

1. Juste enlever ceux qui avaient >3 fois le même jour pour éviter ces cas extrêmes

Autre porblème remarqué bizarre les personnes qui s’abonnent 2 fois dans le même mois presque tous les mois ?

2ème étape : On calcule le nombre d’utilisation de chaque ODD par personne et on inverse les colonnes

**TO DO**

* Vérifier la pertinence de nos données
* Faire des première stats pr voir la cohérence
* Voir quelles colonnes rajouter à notre new data set
* Commencer clustering (je suis chauddd)

561 / 11077085

337795 / 11077085