Travail Datawarehouse & ETL

Le but de ce travail est de vous permettre de faire un datawarehouse à partir d’un fichier plat.

# Mise en situation

Quelque part en inde dans la province de Tamil Nadu en Inde une épicière devenue grossiste voit son commerce croître. Cette dernière avait jusqu’alors un seul magasin elle en a aujourd’hui 5 et à commencer la vente sur internet. Mme Pradesh bonne négociante toujours affable a désormais des problèmes de gestion, surtout au niveau de sa vente par correspondance. Elle commence à avoir beaucoup trop de donné pour son bon vieil Excel. Elle ne sait plus vraiment comment exploiter ses données et ne connait plus vraiment les habitudes de ses meilleurs clients. Ces derniers se détournent progressivement de son business et ce problème d’exploitation de données commencent donc à lui couter une partie de sa clientèle.

Son fils qui est plus un gestionnaire qu’un négociant pense avoir une solution pour le petit réseau de d’épicerie fondé par sa mère. Son fils pense que l’entreprise de sa mère a besoin de passé d’Excel à une base de données partiellement normalisée. Il demande à se mère de lui donner un dataset anonymisé de ses meilleurs clients pour voir comment il pourrait faire pour passer d’Excel à une solution plus normalisé. Son fils fait appel à vous, en votre qualité d’expert pour que vous puissiez lui donner votre avis tant sur la faisabilité, que sur la manière de réaliser cette base de données ? Néanmoins il voudrait aussi savoir si son idée de normalisation partielle est une bonne idée. Votre avis pourrait lui apporter certains éléments qui pourrait lui manquer.

# Les données

Le dataset est un fichier plat de type .csv et se présente comme suit :



Il est composé de 12 colonnes :

* OrderId qui la manière unique d’enregistrer une commande.
* Customer Name qui est le faux prénom du client (données anonymisées oblige).
* Gender qui est le genre (sexe) enregistré par le client.
* Birthdate qui est la date de naissance du client.
* Category qui est la catégorie de produits qui a été commandé par cette commande
* Sub Category qui est une catégorie plus fine
* Order date qui est le moment ou la commande a été passée.
* Region qui est en fait le nom l’entrepôt d’où part la marchandise.
* Sales qui est le chiffre d’affaires rapporté par cette commande
* Profit qui est le bénéfice net dégagé par cette commande
* State l’état de destination de cette commande

Les groupes ne sont pas forcément les mêmes et le dataset peut être légèrement différents. À noter que les données ne sont pas toujours propres et il peut y a avoir des erreurs dans ces dernières. Il vous revient de trouver une solution adaptée à ces problèmes.

# Le but du travail

Le but du travail est de vous permettre d’évaluer votre compréhension des aspects théorique et pratique à la fois du cours d’ETL et de datawarehouse.

## Comprendre le business du cas fictif

L’introduction des données ci-dessous ne vous permettra pas forcément de comprendre totalement le business du client en termes de modélisation. Vous êtes en charge avec les données que vous avez d’affiner cette perception pour comprendre de manière fine les différentes dimensions qui serait le plus adapté pour le business.

## Créer un modèle de donnée de type Datawarehouse

Une fois le business compris, il vous revient de faire un modèle de donnée de type datawarehouse. Pour être plus précis on souhaite un modèle qui facilité l’accès des utilisateurs en lecture. Pour se faire nous avons vus 2 modèles de données en cours de datawarehouse. À vous de voir quel modèle serait le plus exact dans le cas présent. J’attends d’avoir un schéma de votre modélisation, soit via un programme tel que draw.io ou canva, soit via paint, soit même à la main (tant que c’est lisible).

## Faire un flux qui permettent d’avoir le résultat voulu

Finalement il vous faudra construire le flux ETL avec SSIS afin d’amener les données de votre fichiers plats dans une base de données de type SQL-Server. Le flux ne supprimera de préférence pas ou peu les données de la table de destination finale.

## Être critique par rapport aux solutions que vous proposez

Vous n’êtes pas juste des représentants techniques. Posez vous ces questions avant de venir à l’examen, est ce que votre modèle à des défauts. Quels sont les limites de votre démarche et pourquoi avez-vous choisi cette voie la plutôt qu’une autre.