

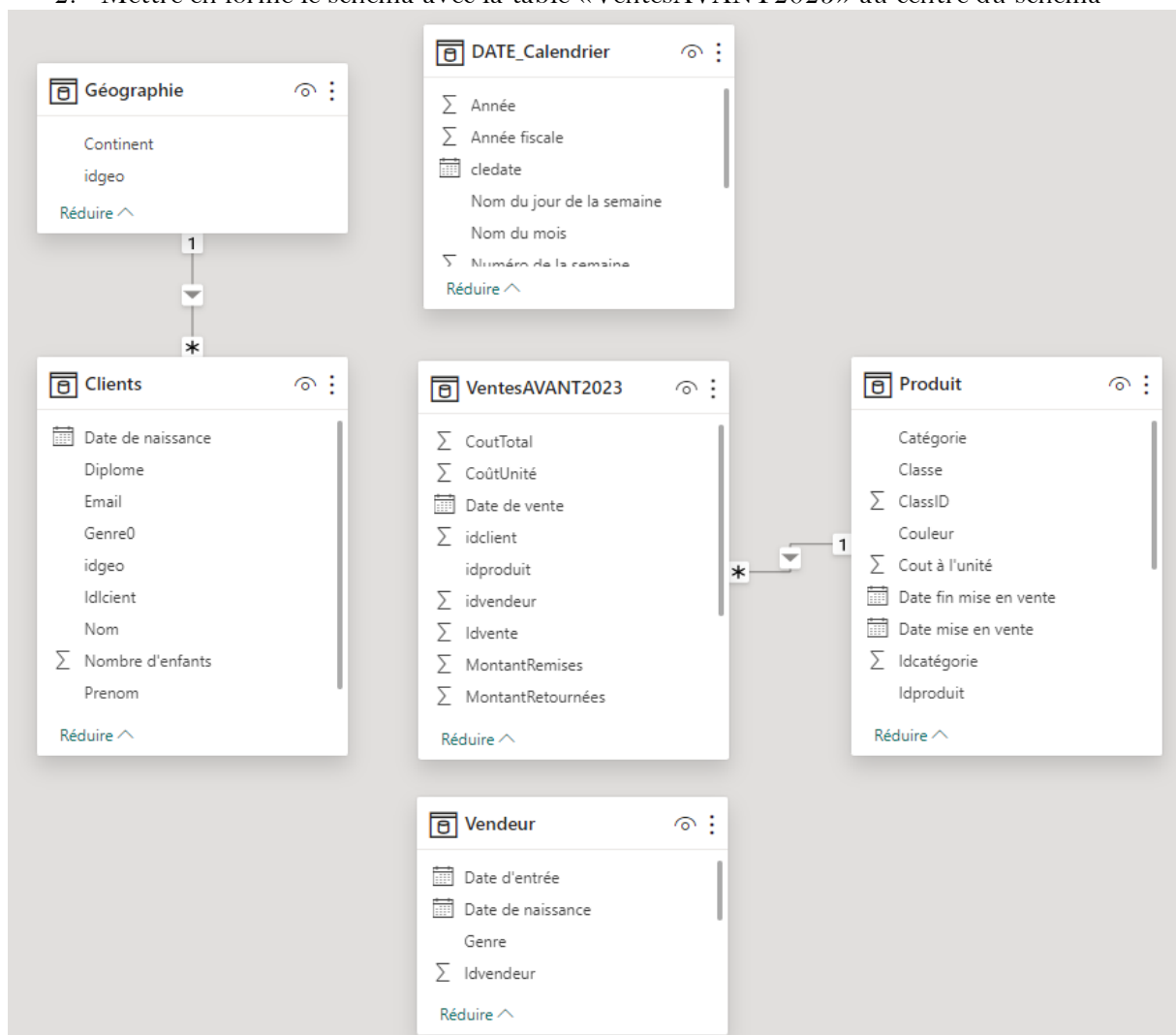
I. Objectifs : Extraction des données

Pour mes tableaux de bords, j'ai besoin de modéliser un schéma qui comportera ces différentes sources de données :

1. « **Ventes et géographie** » → Document Excel contenant 5 feuilles à importer : « Géographie », « Géographie2 », « VentesAVANT2023 », « VentesDEPUIS2023 » et « DATE_Calendrier ».
2. « **Produits** » → document Excel contenant 1 feuille MAIS il faut importer le tableau intitulé Produit ».
3. « **Clients** » → fichier texte à importer,
4. « **Vendeur** » : Connexion Web → Aller chercher le tableau sur la page du site internet suivant : <http://berenguer-formation-conseil.fr/tableau-vendeur-pbi/>
Le tableau sera chargé sous le nom de table1. Il faut le renommer en « Vendeur ».

Fermer la fenêtre Power Query et aller sur l'onglet de gauche « MODELE » dans la fenêtre Power BI

1. Mettre de côté les tables «VentesDEPUIS2023» et « Géographie2 » et SUPPRIMER tous les liens qui existent entre ces deux tables et le reste du schéma,
2. Mettre en forme le schéma avec la table «VentesAVANT2023» au centre du schéma



3. Gérer les relations : Vérifier celles que Power BI a établi automatiquement et Créer les autres. Chaque table possède une clé primaire qui commence par le nom de la table et précédé du mot id. Si on retrouve cet Identifiant dans une autre table, c'est que l'on peut faire une relation par cette clé. Le seul cas où ce n'est pas le cas c'est avec la table *Date_calendrier* (cledate).

Sur les dernières versions, il va manquer trois relations.

- de la table *ventesavant2023* vers la table *table_calendrier* avec les clés Date de vente et Cledate,
- de la table *client* vers la table *ventesavant2023* avec les clés idclient,
- de la table *vendeur* vers la table *ventesavant2023* avec les clés idvendeur.

Il faut se demander pourquoi il ne les a pas faites ?

- Pas le même nom de colonne ?
- Pas le même type de données ?
- Pas de cardinalité 1-N traditionnel.

Ne pas forcer la relation vers la table vendeur tant que les doublons ne sont pas supprimés.

Si on clique sur l'icône « Gestion des relations » sur la barre d'outils du haut, on retrouve un récapitulatif de toutes les relations existantes :

Gérer les relations

Active	De : Table (Colonne)	À : Table (Colonne)
<input checked="" type="checkbox"/>	Clients (idgeo)	Géographie (idgeo)
<input checked="" type="checkbox"/>	VentesAVANT2023 (Date de vente)	DATE_Calendrier (cledate)
<input checked="" type="checkbox"/>	VentesAVANT2023 (idclient)	Clients (Idclient)
<input checked="" type="checkbox"/>	VentesAVANT2023 (idproduit)	Produit (Idproduit)

4. Je teste mes relations dans un tableau croisé dynamique, appelé Matrice dans Power BI.

Exemples de croisement dans le tableau croisé dynamique :

- Faire un tableau croisé dynamique avec « MontantVentes » de la table « VentesAVANT2023 » en valeurs et le champ « classe » de la table « produit » en ligne.
- Faire un tableau croisé dynamique avec « MontantVentes » de la table « VentesAVANT2023 » en valeurs et le champ « année » de la table « Date_Calendrier » en ligne. Ventes pour lesquelles pas d'année ? c'est normal car il y a eu des ventes avant 2019 et que la table date_calendrier ne commence qu'en 2019 😊
- Faire un tableau croisé dynamique avec « MontantVentes » de la table « VentesAVANT2023 » en valeurs et le champ « situation familiale » de la table « client » en ligne. Pas de résultats ? Pourquoi ?

CORRECTION AVEC LA FORMATRICE

II. Objectifs : Préparation des données dans Power Query (si couleur bleu, c'est qu'il faut fermer et appliquer la fenêtre Power Query pour retourner sur la fenêtre power bi)

Lors de la première importation de vos données en entreprise, il faudra impérativement vérifier que :

- Les types de données attribués par Power Query ne sont pas forcément bons car dépendent des 200 premières lignes,
- Que toutes les colonnes sont nécessaires ? sinon, les supprimer !
- Que toutes les lignes sont propres ? sinon, les nettoyer ou les supprimer si besoin de filtres !

Il faut se questionner sur comment optimiser le jeu de données pour son utilisation dans les visuels.

1. Renommer les colonnes « Nom » dans les *tables* « *client* », « *vendeur* » avec la bonne valeur : « nom du client », « nom du vendeur ».
2. Je souhaite remplacer les valeurs vides (dû au CSV) par des valeurs Null reconnus par Power BI dans les colonnes « genre » et « situation familiale » de la *table client* : Je sélectionne avec CTRL les deux colonnes puis je fais un clic droit « Remplacer les valeurs » :

Remplacer les valeurs

Remplacez une valeur dans les colonnes sélectionnées par une autre.

Valeur à rechercher

Remplacer par

Options avancées

Je remplace : « rien » par « null » en minuscule 😊

3. Fractionner la colonne « Pays et ville » dans la *table géographie2*, Attention, lors du fractionnement, il faut utiliser un fractionnement personnalisé afin d'utiliser la virgule et les espaces avant et après elle comme séparateur complet. (Sinon le pays va être écrit avec un espace après et la ville avec espace avant).
Puis, renommer les colonnes proprement

4. Dans la *table client*, supprimer les lettres qui apparaissent dans la colonne idclient (fractionnement spécifique) et renommer la colonne idclient proprement (sans le 1).
5. Supprimer les doublons sur la colonne « idvendeur » de la *table vendeur* afin de n'avoir qu'une seule ligne par vendeur différente.
6. Filtrer les lignes sur la *table* « *VentesAVANT2023* » : sur la colonne « Date de ventes », il faut seulement les ventes à partir de 2019.

SUITE POWER QUERY

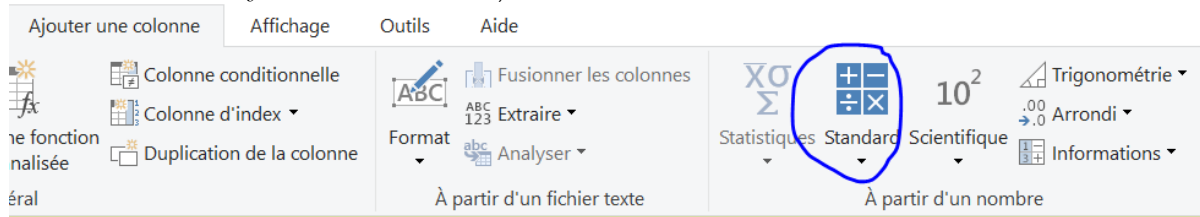
7. Concaténer les lignes de la *table VentesAVANT2023* avec celles de la *table VentesDEPUIS2023* : en utilisant « Ajouter des requêtes ». Ce qui permettra de récupérer les lignes de *VentesDEPUIS2023* dans la *table VentesAVANT2023*. → Il faut se positionner sur la table où l'on souhaite ajouter des lignes, donc bien cliquer sur la *table VentesAVANT2023* puis utiliser « Ajouter des requêtes ».

Renommer ensuite la table *VentesAVANT2023* : *Ventes* !

8. Fusionner la *table geographie* avec la *table geographie2* afin de pouvoir récupérer toutes les colonnes de la *table geographie2* dans la *table géographie* : en utilisant « Fusionner des requêtes » → Pas besoin de récupérer le champ idgeo de la *table geographie2* (car existe déjà dans la *table géographie*).
9. Créer groupes de requêtes : créer groupe « intermédiaire » avec les requêtes *geographie2* et *VentesDEPUIS2023*. Cliquer droit sur la requête : mettre dans autre groupe – créer.
10. Cliquez-droit sur les requêtes dans le groupe intermédiaire et cliquer sur « Activer le chargement » (qui est par défaut coché) ce qui permettra de désactiver le chargement de ses tables dans Power BI. → Car nous n'avons plus besoin de la *table geographie2* ni de la *table VentesDEPUIS2023* car leurs données sont incorporées dans *géographie* et dans *Ventes*.
11. Fermer et appliquer Power Query. Revenir sur l'onglet de gauche « vue de modèle » de Power BI et faire la relation entre la *table ventes* et *table vendeur* via l'idvendeur. Vérifier que la relation entre la table Client et Ventes n'a pas disparu (si c'est le cas, idproduit s'intitule toujours idproduit.1 donc il faut bien renommer cette colonne dans Power Query).

SUITE POWER QUERY

12. Sur la *table Ventes* : Créer une colonne qui nous permette de calculer les quantités réellement vendues : il faut donc soustraire les quantités retournées aux quantités vendues (Onglet du ruban « Ajouter une colonne ») :



13. Sur la *table Ventes*. Créer une colonne personnalisée qui nous permette de calculer la part de bénéfices par ventes. On a besoin donc de soustraire le montant des ventes par celui des coûts totaux avant de diviser par le montant des ventes global.

→ Penser à renommer la colonne, renommer l'étape appliquée à droite et modifier le type de données de votre nouvelle colonne.

$$=([MontantVentes] - [CoûtTotal]) / [MontantVentes]$$

14. Sur la *table Produit*: Dans l'onglet, « ajouter une colonne », fusionner les colonnes Idproduit et nom de produit en les séparant par un point-virgule.

15. Sur la *table clients* : Créer une nouvelle colonne Genre (à partir de la colonne genre 0, avec Femme à la place de F et Homme à la place de H) en utilisant une colonne conditionnelle.

→ Penser à renommer la colonne, renommer l'étape appliquée à droite et modifier le type de données de votre nouvelle colonne.

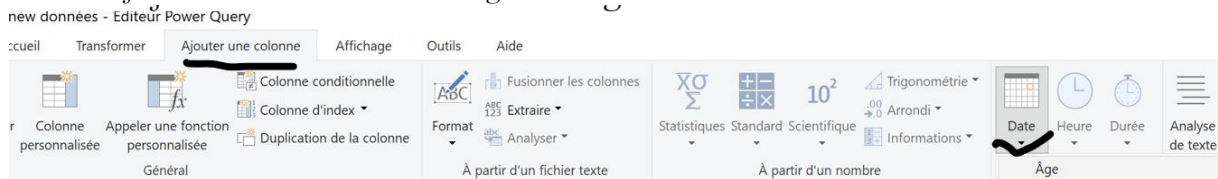
16. Sur la *table clients* : Créer une nouvelle colonne Situation familiale (à partir de la colonne situation familiale0, avec Célibataire à la place de C, Mariée à la place de M et Pacsée à la place de P) en utilisant une colonne conditionnelle.

→ Penser à renommer la colonne, renommer l'étape appliquée à droite et modifier le type de données de votre nouvelle colonne.

17. Concaténer les deux nouvelles colonnes Genre et situation familiale ensemble dans une nouvelle colonne en fusionnant les colonnes avec un espace en guise de séparateur !

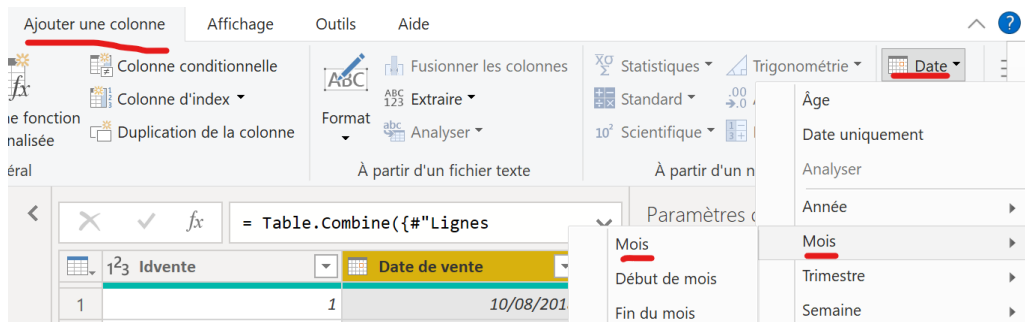
18. Création de l'âge du client en 3 étapes (une seule création de colonne puis transformations dessus) :

- Ajouter une colonne : Date = Age



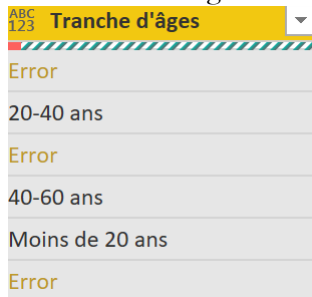
- Transformer : Durée – Nombre d'années
- Transformer : Arrondir à l'entier inférieur

19. Création du numéro du mois dans la *table Ventes* en utilisant le champ *Date de vente*. Ajouter une colonne, puis cliquer sur Date, puis sur Mois et Mois (numéro du mois).



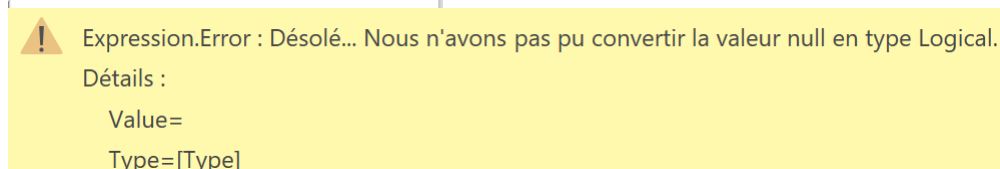
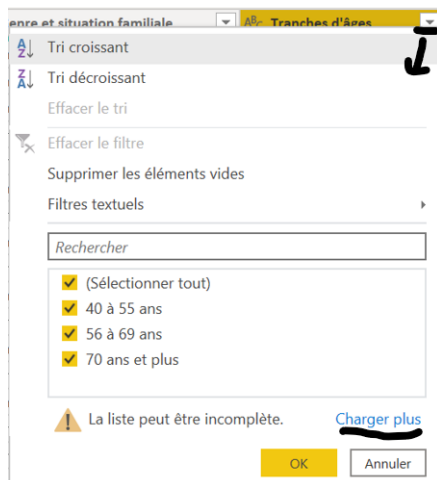
20. *Seulement pour ceux ayant déjà terminé* : Création des tranches d'âges des **CLIENTS** = Moins de 20 ans, 20-39 ans (moins de 40), 40 à 59 ans, 60 ans et plus.

Attention message d'erreur possible ou erreurs qui s'affichent d'office dans les données.



Depuis version août 2022 : cliquer sur la cellule à droite du mot error pour afficher le message d'erreur.

Sur les anciennes versions, utiliser la petite flèche blanche à droite de l'entête de colonne créée et faire « Charger plus ». Les erreurs et l'explication s'affichera à gauche au-dessus du bouton OK.



Il est nécessaire de modifier l'erreur.

2 possibilités :

- *Remplacer erreurs (en cliquant droit sur la colonne en question puis dire par quelle valeur remplacer toutes vos erreurs),*
- *Retourner sur l'étape appliquée précédente (colonne conditionnelle) et rajouter la notion des valeurs Null dans la colonne conditionnelle en tout premier comme ce qui suit :*

1 Cliquer sur ajouter une clause

2 Cliquer tout à droite, sur les 3 petits points de la ligne « Plus » et « Déplacer vers le haut » afin de faire remonter cette nouvelle clause tout en haut

3 Remplir condition : Age est égal à null et là il les gère bien

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ①
Si	Âge	égal à	ABC 123 null	Alors	ABC 123 NC
Autre...	Âge	est inférieur à	ABC 123 20	Alors	ABC 123 Moins de 20 ans

21. Seulement pour les plus avancés et ayant déjà vu du code SQL ou autre :

Tenter de créer directement une nouvelle colonne (via colonne personnalisée) qui aura les valeurs « Homme marié », « Femme mariée » etc. mais sans créer/utiliser les colonnes conditionnelles en amont...

Juste du code → si H pour la colonne Genre et si C pour la colonne Situation familiale alors Homme célibataire...

La syntaxe : Si dans la colonne champ1, la valeur est A1 ET dans la colonne champ2 la valeur est B ALORS je dois créer la valeur « valeur AB » dans ma nouvelle colonne

if [champ1] = "valA" and [champ2] = "B" then "valeur AB"

Else if [champ1] = "valB" and [champ2] = "B" then "valeur BB"

Else if etc.

Else "blabla"

```

=if [Genre0] = "H" and [Situation familiale0] = "M" then "Homme Marié"
else if [Genre0] = "H" and [Situation familiale0] = "C" then "Homme célibataire"
else if [Genre0] = "H" and [Situation familiale0] = "P" then "Homme pacsé"
else if [Genre0] = "F" and [Situation familiale0] = "M" then "Femme mariée"
else if [Genre0] = "F" and [Situation familiale0] = "C" then "Femme célibataire"
else if [Genre0] = "F" and [Situation familiale0] = "P" then "Femme pacsée"
else "NC"

```

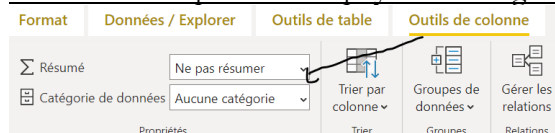
CORRECTION AVEC LA FORMATRICE

III. Fenêtre Power BI – Partie Cube

1. Catégoriser les données : Il faut être positionné sur l'onglet de gauche **Données ou rapport** dans Power BI. Puis dans le ruban en haut, « Outils de colonne » dès que l'on clique sur un champ.

■ Catégoriser les colonnes de la table géographie : localité pour champ ville et région pour Etat, etc.

Est-ce un code postal ? Un pays ? Une région, etc.










2. Définir Année de la table *Date_calendrier* comme « Ne pas résumer » dans l'onglet 'Outils de colonnes' car on ne souhaite pas qu'il le mette en valeurs dans les visuels mais bien qu'il l'affiche en lignes/colonnes/axes/légende lors du cochage de la colonne.
3. Définir table des *Date_calendrier* (Onglet de gauche Données ou rapport) : cliquer droit sur la table à droite : *table_calendrier* et sélectionner « Marquer comme table de dates » puis sélectionner le champ Date comme Colonne de date sur la petite fenêtre qui s'affiche ensuite.

4. Trier la colonne « nom du mois » en fonction du numéro de mois.

→ Onglet « Outils de colonnes » quand on clique sur le champ en question puis « Trier par colonne »

5. Créer une hiérarchie dans la table *Date calendrier* et une autre table géographie

▼ <input type="checkbox"/>  Année Hiérarchie	▼ <input type="checkbox"/>  Continent Hiérarchie
<input type="checkbox"/> Année	<input type="checkbox"/>  Continent
<input type="checkbox"/> Trimestre	<input type="checkbox"/>  Pays
<input type="checkbox"/> Nom du mois	<input type="checkbox"/>  Etat ou région
<input type="checkbox"/>  cledate	<input type="checkbox"/>  Ville

Une hiérarchie peut être créée automatiquement au chargement d'une table qui ressemble pour lui à une table de date ou quand on crée une table de date.

Si elle n'existe pas, il faudra aller sur le modèle de données, cliquer droit sur champs et créer une Hiérarchie, en glissant tous les champs à l'intérieur.

6. Tester ensuite vos hiérarchies dans un tableau et un graphique.