



# Rapport de traitement des données

Titre du projet	
	Collectez des données en respectant les normes RGPD

Version	Auteur	Description	Date
V1	Aurelia DE INFANTI	Rapport	22.01.2025

## Introduction

**Contexte** : La société a été sanctionnée par la CNIL d'une limitation temporaire des traitements et doit s'assurer que les règles du RGPD soient appliquées et intégrées afin de demander la levée de la sanction. Pour ce faire, des actions sont à mener.

### Objectifs du projet

- **Conformité réglementaire** : Assurer que toutes les données soient gérées conformément aux exigences du RGPD.
- **Optimisation des processus** : Mettre en place des flux de traitement des données qui soient sécurisés et transparents.
- **Valorisation des données** : Faciliter l'analyse des données tout en garantissant leur anonymisation.

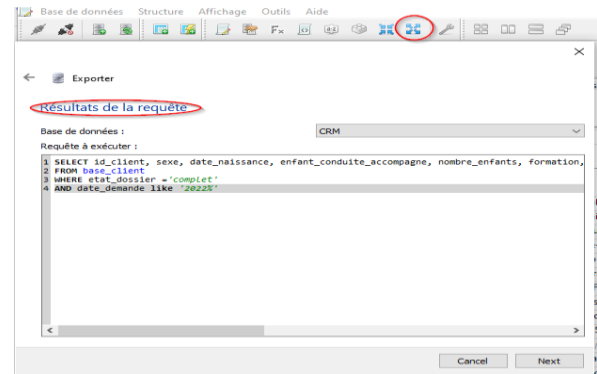
L'objectif de cette requête est d'extraire des informations spécifiques depuis la table « **base\_client** » pour les clients : Dont la **date de demande** est dans l'année **2022** et dont l'**état du dossier** est marqué comme « **complet** ». Empêcher que l'équipe de performance commerciale n'ait accès aux données tarifaires détaillées.

## Contenu du rapport :

### 1. Extraction des données

#### A. La requête :

```
SELECT id_client, sexe, date_naissance,
enfant_conduite_accompagne, nombre_enfants,
formation, adresse, usage_vehicule, type_vehicule,
age_vehicule, type_conduite, formule, tarif_devis,
revenus, date_demande, est_rouge, points_perdus
FROM base_client
WHERE etat_dossier = 'complet'
AND date_demande like '2022%'
```



Cette sélection permet de:

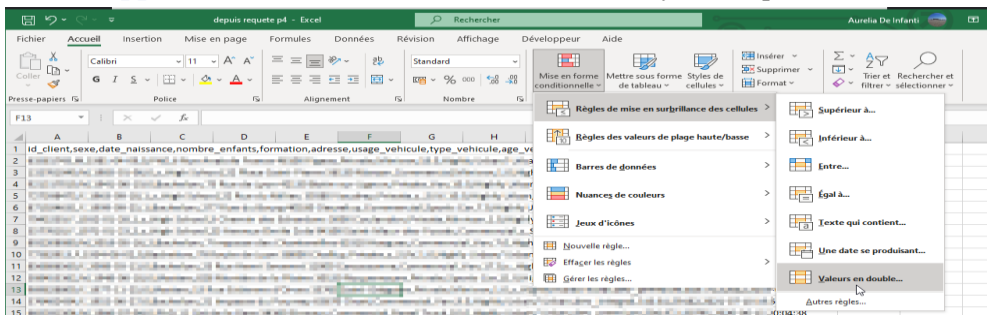
Identifier uniquement les clients de manière anonyme

Analyser les caractéristiques démographiques et socio-économiques sans révéler d'informations personnelles sensibles

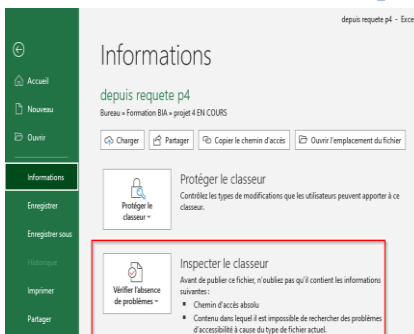
Étudier les comportements et les préférences en matière d'assurance automobile

Évaluer la valeur client et les opportunités de vente croisée

#### B. Supprimer les doublons dans fichier csv : il n'y en n'a pas



#### C. Vérifier l'absence de problème : il n'y en n'a pas

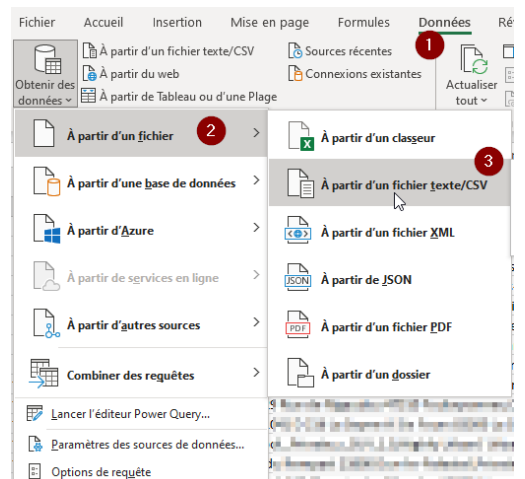


#### D. Ouverture POWER QUERY

Sélectionner le fichier : Naviguer jusqu'à l'emplacement du fichier et sélectionnez le.

Prévisualisation des données : Une fenêtre s'ouvre avec un aperçu des données. Vérifier que le délimiteur (par exemple, une virgule) est correctement détecté. Cliquez sur

Transformer les données.

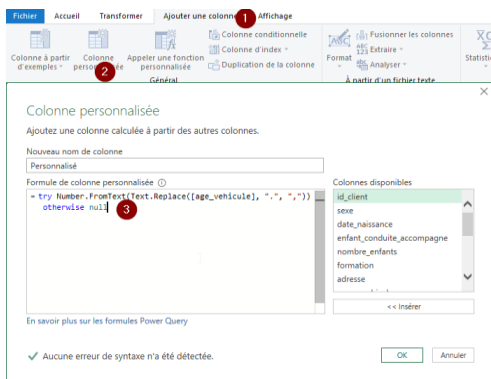


## 2. Modifier type de données

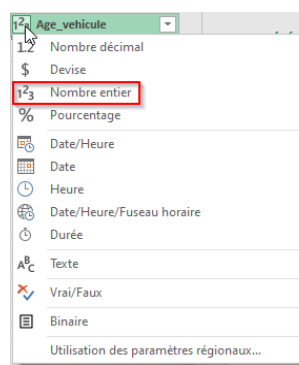
**A. Age du véhicule** : type de donnée : texte. A modifier en nombre

Remplacer les points par des virgules :

Dans le ruban, cliquez sur **Ajouter une colonne** > **Colonne personnalisée** = try  
*Number.FromText(Text.Replace([age\_vehicule], ".", ",")) otherwise null*



Renommer la colonne personnalisée et supprimer les colonnes inutiles



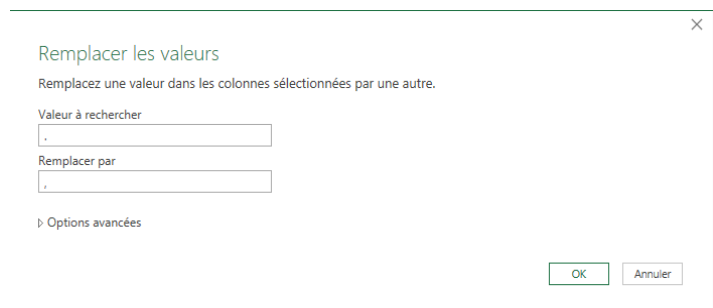
Modifier le type de données en nombre entier

**B. Tarifs devis** : Format texte à convertir en nombre

Idem (voir **A. Age du véhicule**) ou :

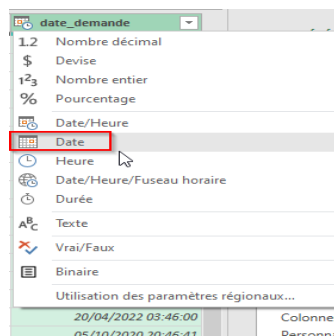
**Utiliser l'option Remplacer des valeurs :**

- Sélectionnez la colonne « tarif\_devis ».
- Clic droit : **Remplacer des valeurs**.
- Valeur à rechercher : «. » (Le point).
- Valeur de remplacement : «, » (La virgule).



**C. Date de la demande** : type de donnée Date/Heure à modifier en Date

- Clic droit puis choisir **Date**



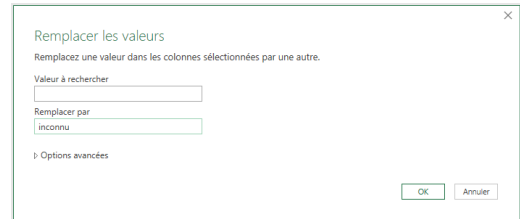
### 3. Repérage des données vides/ null et les renseigner à « inconnu »

#### A. Adresse : 22 lignes vides = remplacées par « Inconnu »

Sélectionnez la colonne « Adresse ». Clic droit : **Remplacer des valeurs.**

- Valeur à rechercher : «.» (vide).
- Valeur de remplacement : « inconnue »

*Si format date et nombre : les données ont été remplacées par null lors du changement de type de données.*



#### B. revenus : Idem, On remplace les valeurs « 0 » par null

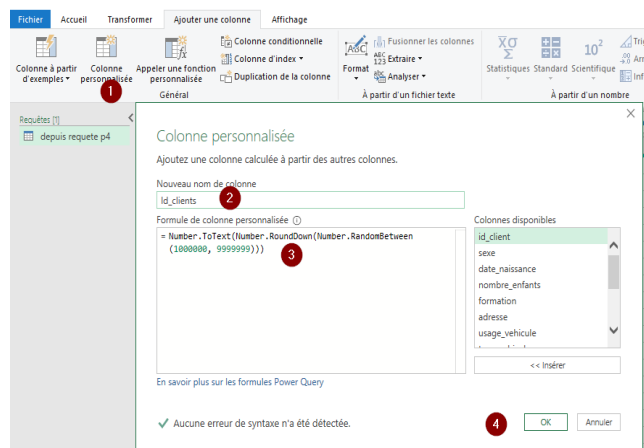
### 4. Anonymiser depuis Power Query :

#### A. id\_client : à anonymisé

- Dans le ruban, cliquer sur **Ajouter une colonne > colonne personnalisée.**
- Nommer la colonne, entrer la formule :

**= Number.ToText(Number.RoundDown  
(Number.RandomBetween(1000000, 9999999)))**

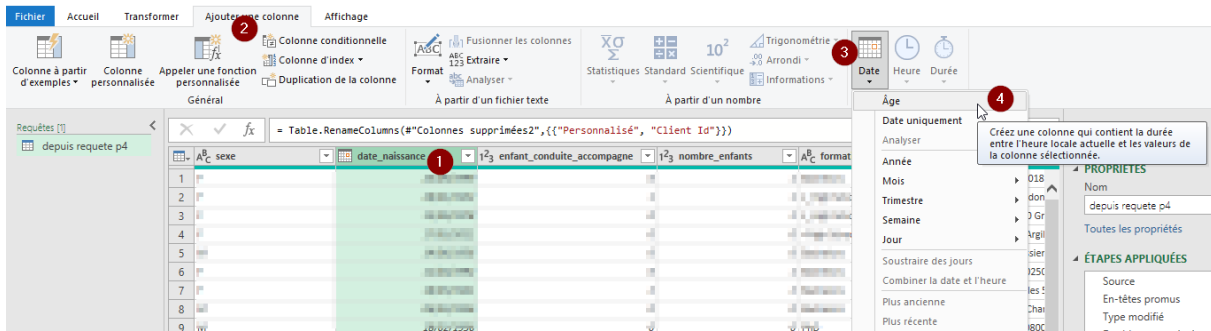
Puis suppression de la colonne devenue inutile



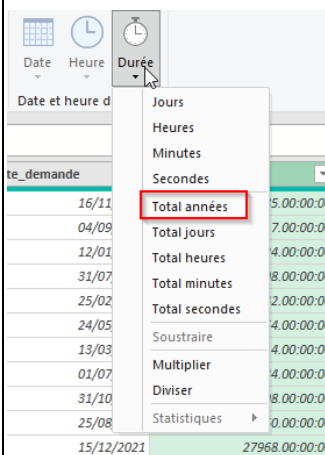
## B. Date\_naissance : Généraliser les dates de naissance (par tranche âge)

### Ajouter une colonne d'âge :

- Sélectionnez la colonne de dates de naissance.
- Dans le ruban, cliquer sur **Ajouter une colonne > Date > Age**.



- Cliquer sur durée > **Total années**



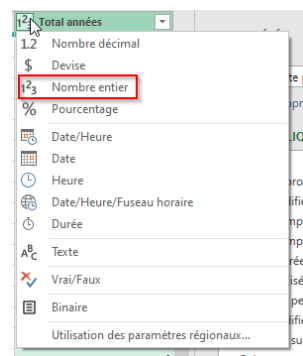
- Puis choisir **Nombre entier**

Supprimer les colonnes inutiles (date naissance et Age)

Nous avons alors les ages des clients, il faut effectuer les **tranches d'âges**

*Il n'est pas possible d'ajouter une colonne à partir d'exemple car nous voulons faire ressortir les moins de 18 ans*

*Ex : pas de cohérence avec les personnes qui ont déclaré une date erronée (7 apparait dans 4-13 ans et 17 dans 13-22 ans).*

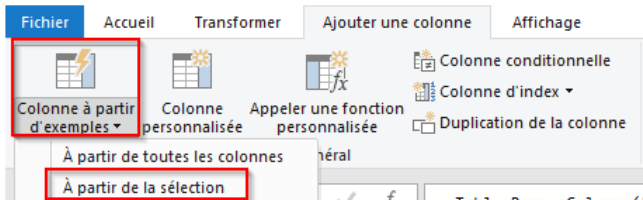


- Dans le ruban, cliquer sur **Ajouter une colonne > colonne personnalisée**
- Nommer la colonne, entrer la formule :  

```
if [Total années]< 18 then "Moins de 18 ans"
else if [Total années] >= 18 and [Total années] < 30 then "18-29 ans"
else if [Total années] >= 30 and [Total années] < 40 then "30-39 ans"
else if [Total années] >= 40 and [Total années] < 50 then "40-49 ans"
else if [Total années] >= 50 and [Total années] < 60 then "50-59 ans"
else if [Total années] >= 60 and [Total années] < 70 then "60-69 ans"
else if [Total années] >= 70 and [Total années] < 80 then "70-79 ans"
else "80 et plus"
```
- Supprimer la colonne inutile (Total années)

**C. nombre\_enfants** (oui/non) : colonne à partir d'exemple : 0 =non, chiffre = oui

- Dans le ruban, cliquer sur **Ajouter une colonne** > **colonne à partir d'exemple** > A partir de la sélection > entrer des exemples



123	nombre_enfants	✓	A <sup>B</sup> C formation	Personnalisé
	1			oui
	1			oui
	2			oui
	0			non
	0			non
	1			oui
	2			oui
	0			non
	0			non
	0			non
	2			oui

ou ajouter une **colonne conditionnelle**

- Dans le ruban, cliquer sur **Ajouter une colonne** > **colonne conditionnelle** > renommer la colonne > complétez l'opérateur, valeur, sortie > OK



- Supprimer la colonne inutile (nombre d'enfant)

**D. revenus** (catégorisé)

- Dans le ruban, cliquer sur **Ajouter une colonne** > **colonne à partir d'exemple** > A partir de la sélection > entrer des exemples
- Format texte : null est remplacé par inconnu

Renommer puis supprimer la colonne inutile (revenus)

123	revenus	✓	Plage
	125001		120000-129999
	50000		50000-59999
	21000		20000-29999
	null		null
	null		null
	79000		70000-79999
	130000		130000-139999
	51000		50000-59999
	115000		110000-119999
	null		null
	24000		20000-29999
	20000		20000-29999
	74000		70000-79999
	null		null
	38000		30000-39999
	45000		40000-49999
	108000		100000-109999
	53000		50000-59999
	76000		70000-79999

D. revenus

**E. Conduite accompagnée** (oui/non)

Idem : **colonne à partir d'exemple** : 0 =non, chiffre = oui

Renommer puis supprimer la colonne inutile (enfant\_conduite accompagnée)

**F. Est rouge** (oui/non)

Idem : **colonne à partir d'exemple** : no =non, yes = oui

Renommer puis supprimer la colonne inutile (est\_rouge)

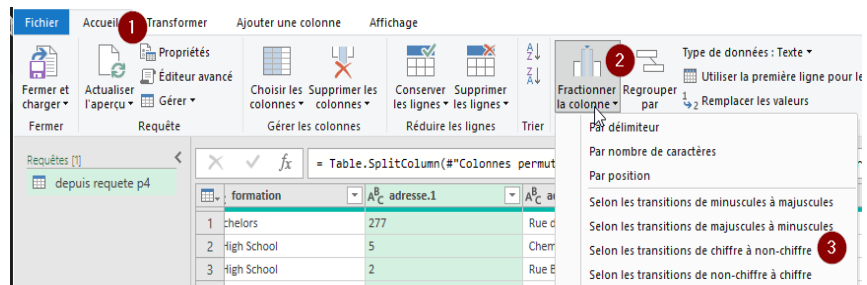
Personnalisé	Plage
	280-300
	280-300
	280-300
	360-380
	320-340
	280-300
	320-340
	240-260
	320-340
	320-340
	440-460
	300-320
	500-520
	240-260
	300-320
	280-300
	280-300
	480-500
	540-560
	300-320
	500-520
	440-460
	240-260
	640-660

#### G. tarif devis (catégorisé)

- Dans le ruban, cliquer sur **Ajouter une colonne** > **colonne à partir d'exemple** > A partir de la sélection > entrer des exemples
- Ajout de **colonne à partir d'exemple**, je fais catégorie par 20  
Renommer et supprimer la colonne inutile (tarifs\_devis)

#### H. Adresse (pour code postal)

- Dans le ruban, cliquer sur **Accueil** > **Fractionner la colonne** > Selon les transitions de chiffre à non-chiffre



- On ne garde que la colonne « code postal »

*Qu'on n'est pas obligé de convertir en chiffre, car pas de calcul à faire avec code postal et il y a du texte dedans*

A <sub>1</sub> C <sub>1</sub> adresse.1	A <sub>1</sub> C <sub>2</sub> adresse.2	A <sub>1</sub> C <sub>3</sub> adresse.3	A <sub>1</sub> C <sub>4</sub> adresse.4	A <sub>1</sub> C <sub>5</sub> code postal
277	Rue du Toussaint (03000)	Chapelle		40180
5	Chemin de la Chapelle (03000)	Chapelle		03000
2	Rue de la Chapelle (03000)	Chapelle		04800
66	Parc de la Chapelle (03000)	Chapelle		06370
19	Rue de la Chapelle (03000)	Chapelle		19100
7	Rue de la Chapelle (03000)	Chapelle		70250
87	Rue de la Chapelle (03000)	Chapelle		58000
10	Rue de la Chapelle (03000)	Chapelle		73460
30	Quai de la Chapelle (03000)	Chapelle		30800
29	Rue de la Chapelle (03000)	Chapelle		47510
1041	D 216	La Seyne-sur-Mer		83500
Inconnue				inconnu
6	Rue de la Chapelle (03000)	Rue de la Chapelle		23800

## 5. Détail de l'analyse d'extraction et d'anonymisation des données

### A. Depuis extraction des données, nous avons retenu :

- **id\_client** : à **anonymisé** pour avoir un id unique
- **sexe** : pour caractéristiques démographiques et socio-économiques
- **date\_naissance** : à **catégoriser** par tranche d'âge pour caractéristiques démographiques et socio-économiques
- **adresse** : **ne garder que code\_postal**, caractéristiques démographiques et socio-économiques
- **revenus** caractéristiques démographiques et socio-économiques
- **nombre\_enfants** : à **catégorisé** : oui/non, caractéristiques démographiques et socio-économiques
- **enfant conduite accompagné** : à **catégorisé** : oui/non, caractéristiques démographiques et socio-économiques
- **formation** : caractéristiques démographiques et socio-économiques
- **usage\_vehicule** : comportements et les préférences en matière d'assurance automobile
- **type\_vehicule** : comportements et les préférences en matière d'assurance automobile
- **age\_vehicule** : comportements et les préférences en matière d'assurance automobile
- **type\_conduite** : comportements et les préférences en matière d'assurance automobile
- **formule** : valeur client et les opportunités de vente croisée
- **tarif\_devis** : à **catégorisé** par tranche, valeur client et les opportunités de vente croisée
- **date\_demande** : 2022, demandé par le métier
- **est\_rouge** : utile car + accident si véhicule rouge, valeur client et les opportunités de vente croisée
- **points\_perdus** : demandés par certaines assurance, valeur client et les opportunités de vente croisée
- **etat\_dossier** : complet, demandé par le métier

### B. Nous n'avons pas retenu

- **metier** : trop précis, demander le secteur de métier pour les prochaines récoltes de données
- **employeur** : inutile et trop précis
- **num ss** : donnée très sensible et non utile
- **groupe sanguin** : donnée très sensible et inutile
- **id\_site web** : inutile et trop précis
- **nom** : inutile et trop précis
- **email** : inutile et trop précis
- **valeur résidence principale** : trop précis, inutile
- **lat – lon** : trop précis

### C. Puis avons analysé et anonymisé si besoin

- **id\_client** : **anonymisé**
- **sexe** : pas besoin de modifier
- **date\_naissance** : **catégorisé** par tranche d'âge
- **adresse** : **devient** code\_postal
- **revenus** : **catégorisé** par tranche
- **nombre\_enfants** : **catégorisé** : oui/non
- **enfant conduite accompagné** : **catégorisé** : oui/non
- **formation** : pas besoin de modifier
- **usage\_vehicule** : pas besoin de modifier
- **type\_vehicule** : pas besoin de modifier
- **age\_vehicule** : pas besoin de modifier
- **type\_conduite** : pas besoin de modifier
- **formule** : pas besoin de modifier
- **tarif\_devis** : **catégorisé** par tranche
- **date\_demande** : **retrait** de l'heure
- **est\_rouge** : **modifié** en français
- **points\_perdus** : pas besoin de modifier
- **etat\_dossier** : complet



## 6. Edition du dictionnaire des données

Nom Champ	Type	Description	infos
sexe	string	sexe	
id_client	string	identifiant unique du client dans notre CRM	A été modifié pour indexation
formation	string	type de formation du client	
usage_vehicule	string	usage principal du véhicule	
type_vehicule	string	type de véhicule	
est_rouge	string	indique si le véhicule du client est rouge	A été modifié : yes=où, no=non
points_perdus	nb entier	points retirés sur le permis de conduire à la date de la demande d'assurance	
age_vehicule	nb entier	âge du véhicule à assurer	
type_conduite	string	type de conduite envisagée	
date_demande	date	date de la demande de devis faite sur internet	
formule	string	formule choisie par le client lors de sa demande internet	
Tranche d'âge	string	Tranche d'âge.	!!clients de - 18 ans, erreur lors de l'intégration des données
Enfants	string	1 ou plusieurs enfants = oui, pas d'enfant : non	
conduite accompagnee	string	1 ou plusieurs enfants en conduite accompagnée = oui, pas d'enfant : non	
tranche revenus	string	Tranche de revenu de référence du client	!! certaines données sont null
tranche tarifs_devis	string	Tranche tarif accordé au client en fonction de son profil	
code postal	string	code postal du client	

### Conclusion

Ce rapport souligne les adaptations nécessaires pour assurer la conformité au RGPD tout en optimisant l'utilisation des données personnelles.

Les principes clés du RGPD, tels que la minimisation des données, la finalité légitime, la transparence et la sécurité, ont été appliqués tout au long du processus, de la collecte à l'anonymisation partielle.

Le fichier de données créé est spécifiquement conçu pour répondre aux besoins de l'équipe de performance commerciale, en tenant compte de leurs exigences opérationnelles et analytiques

Il est important de noter que d'autres services de l'entreprise pourraient avoir des besoins différents en matière de données, nécessitant des ajustements spécifiques tout en respectant le RGPD.

Pour renforcer la conformité, il est recommandé d'ajouter une colonne calculée dans la base de données pour déterminer l'ancienneté des données.

Cette mesure facilitera la gestion des durées de conservation, un aspect crucial pour une politique d'archivage, de suppression ou d'anonymisation appropriée.

Enfin, il est essentiel de maintenir une vigilance constante et de réaliser des audits réguliers pour adapter les processus aux évolutions technologiques et aux recommandations des autorités compétentes, notamment la CNIL.

Ces efforts contribueront à maintenir la confiance des clients et à renforcer la réputation de l'entreprise en matière de gestion éthique et sécurisée des données.