



# Modèle de rapport de traitement des données

<b>Titre du projet</b>	<i>Mise en conformité RGPD</i>
------------------------	--------------------------------

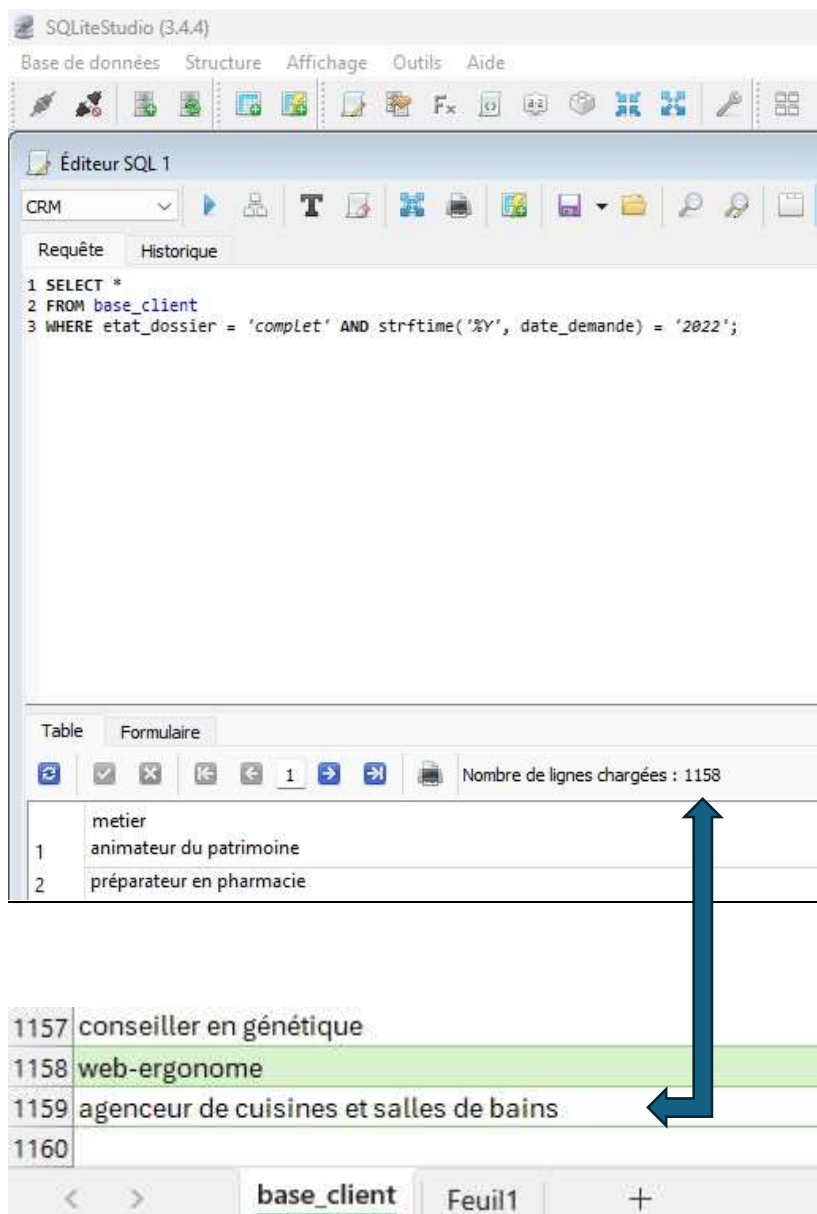
Version	Auteur	Description	Date
V1	Ah-see	Rapport	21/03/2024

Introduction
<p>Un ancien client a porté plainte auprès de la CNIL pour non-respect du règlement général de la protection des données (RGPD).</p> <p>Un devis lui a été transmis avec un malus de 25 % par rapport à un vieux relevé d'informations de plus de 6 ans transmis pour une demande antérieure.</p> <p>La CNIL a mis en demeure Dev'Immédiat de se mettre en conformité par rapport au RGPD avec une suspension temporaire des traitements de données de 6 mois.</p> <p>Une forte sanction financière n'est pas à exclure.</p> <p>La mission consiste à établir un plan d'action pour se mettre en conformité très rapidement par rapport au RGPD.</p>

## Contenu du rapport :

### Etape 1: Requête SQL

Après l'importation de la base de données dans SQLite, nous procédons à l'extraction des données de l'année en cours (la demande d'assurance a été effectuée en 2022) pour les clients dont l'état du dossier est « complet » à partir de la base SQL :



The screenshot shows the SQLiteStudio interface. The SQL editor contains the following query:

```
1 SELECT *
2 FROM base_client
3 WHERE etat_dossier = 'complet' AND strftime('%Y', date_demande) = '2022';
```

Below the editor, the results are displayed in a table view. The table has two columns: an index and a 'metier' (job) column. The results show a list of professions, with row 1158 highlighted in green. A blue arrow points from the 'Nombre de lignes chargées : 1158' status bar to the highlighted row 1158.

	metier
1	animateur du patrimoine
2	préparateur en pharmacie
...	...
1157	conseiller en génétique
1158	web-ergonome
1159	agenceur de cuisines et salles de bains
1160	

base\_client Feuil1

1 158 lignes chargées dans SQLite et nous retrouverons le même nombre de lignes dans l'export Excel avec 1 ligne à en-tête en plus dans Excel:  $1158+1=1\ 159$ .



## Explication de la requête SQLite:

**SELECT \*** : Sélectionne toutes les colonnes de la table.

**FROM base\_client** : Spécifie la table à partir de laquelle les données doivent être extraites.

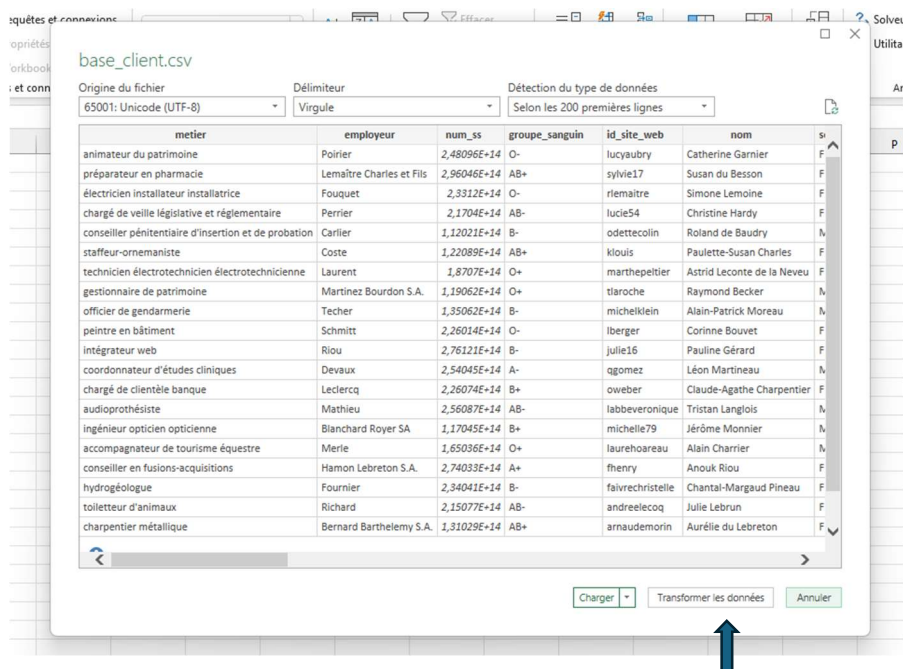
**WHERE etat\_dossier = 'complet'** : Filtre les résultats pour inclure uniquement les lignes où l'état du dossier est "complet".

**AND strftime ('%Y', date\_demande) = '2022'** : Ajoute une condition pour extraire uniquement les données de l'année 2022 en utilisant la fonction strftime pour extraire l'année à partir de la colonne date\_demande.

On exportera ensuite le fichier CSV de la requête de SQLite vers Excel.

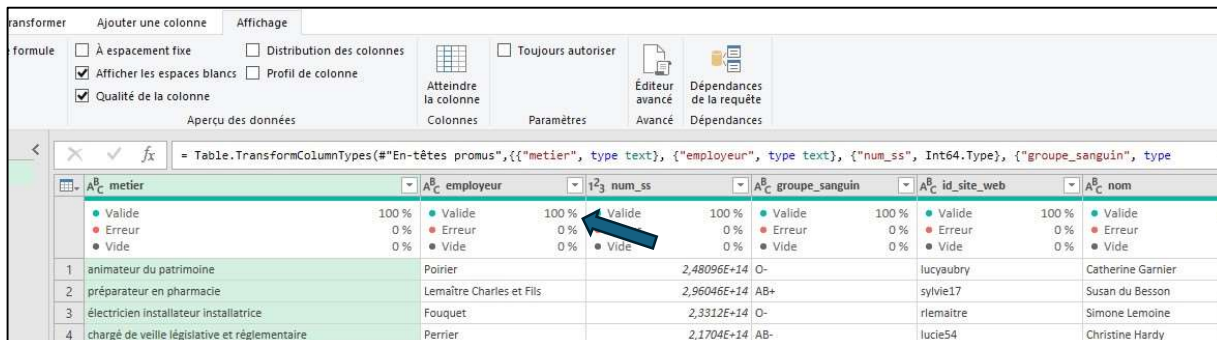
## Etape 2: importation du fichier CSV dans Power Query

Power Query est un logiciel ETL (**E**xtraction **T**ransform **L**oad) pour le traitement des données. Nous importons le fichier CSV dans Power Query.



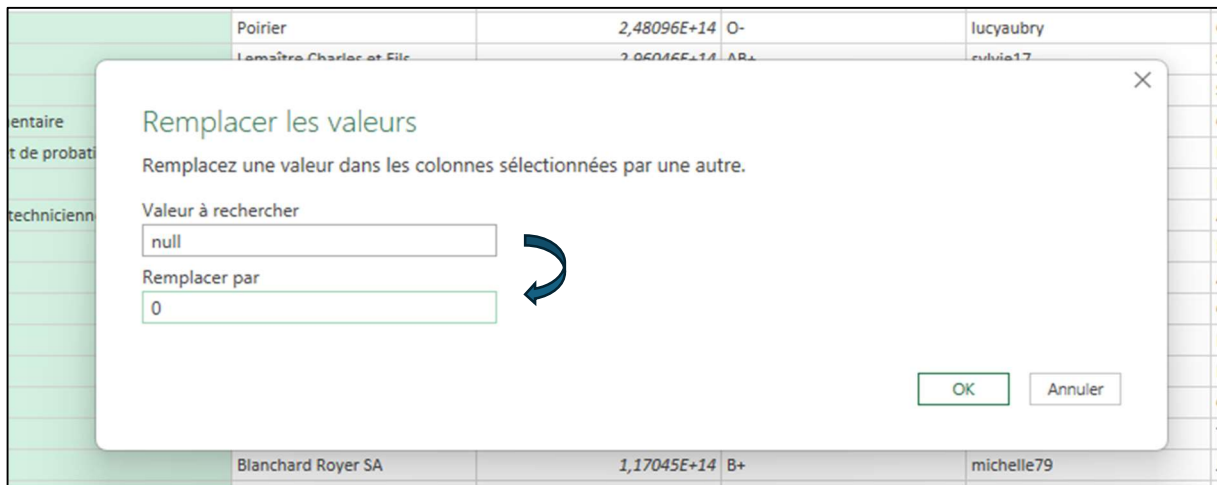
## Etape 3: vérification des erreurs du fichier dans Power Query

Nous vérifions toutes les colonnes du fichier afin d'avoir 100 % Valide, 0 % Erreur et 0 % Vide.



	metier	employeur	num_ss	groupe_sanguin	id_site_web	nom
	100 % Valide 0 % Erreur 0 % Vide	100 % Valide 0 % Erreur 0 % Vide	100 % Valide 0 % Erreur 0 % Vide	100 % Valide 0 % Erreur 0 % Vide	100 % Valide 0 % Erreur 0 % Vide	100 % Valide 0 % Erreur 0 % Vide
1	animateur du patrimoine	Poirier	2,48096E+14	O-	lucyaubry	Catherine Garnier
2	préparateur en pharmacie	Lemaître Charles et Fils	2,96046E+14	AB+	sylvie17	Susan du Besson
3	électricien installateur installatrice	Fouquet	2,3312E+14	O-	rlemaître	Simone Lemoine
4	chargé de veille législative et réglementaire	Perrier	2,1704E+14	AB-	lucie54	Christine Hardy

Nous remplaçons des valeurs null par zéro dans les 2 colonnes (Revenus et Valeur\_residence\_prin) avec le menu Transformer / Remplacer les valeurs



	employeur	num_ss	groupe_sanguin	id_site_web	nom
	Poirier	2,48096E+14	O-	lucyaubry	
	Lemaître Charles et Fils	2,96046E+14	AB+	sylvie17	
	Blanchard Royer SA	1,17045E+14	B+	michelle79	

Nous remplaçons également des espaces vides par 0 dans la colonne Age\_vehicule avec la même fonction (Remplacer les valeurs).

Les 3 colonnes passent à 100 % Valide.

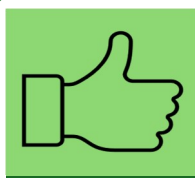
En poursuivant notre contrôle, nous voyons que la colonne Adresse comporte 3 lignes vides...

## **Etape 4: remplissage des données manquantes**

Nous retrouvons les 3 adresses manquantes sur Google Maps à partir de leurs latitudes et leurs longitudes en retournant dans Excel car nous ne pouvons pas modifier ou rajouter les données dans Power Query.

	Z	AA	AB
vis	adresse	lat	lon
	139 Bd de la Madeleine, 06000 Nice	43.702801	7.235774
	20-31 Rue Vincent Barnoin, 06420 Clans	43.995567	7.14744
	10 Av. du Soleil d'Or, 06230 Villefranche-sur-Mer	43.705142	7.305869

**Retournons sur Power Query = nous avons 100 % des colonnes valide**



## **Etape 5: Mise en conformité du fichier dans Power Query**

Le fichier est prêt pour être mis en conformité par rapport au RGPD.

Nous procédons à la minimisation des données.

Le principe de minimisation des données est un principe fondamental du RGPD qui stipule que les organisations ne doivent collecter, traiter et conserver que les données personnelles strictement nécessaires pour atteindre des finalités spécifiques et légitimes.

Ce principe vise à limiter la quantité de données personnelles traitées afin de réduire les risques pour la vie privée des individus et de garantir que les données ne sont pas utilisées de manière excessive ou inappropriée.

Nous commencerons par supprimer les données sensibles qui sont également appelées données à caractère personnel sensibles et sont des informations particulièrement sensibles sur un individu qui nécessitent une protection accrue en vertu du RGPD et d'autres lois sur la confidentialité des données.

Ces données peuvent inclure des informations qui, si elles étaient divulguées ou traitées de manière inappropriée, pourraient causer un préjudice significatif à la personne concernée.

Dans la liste des colonnes du fichier, nous retrouvons une seule donnée sensible qui est une donnée de santé, le groupe sanguin et nous la supprimons.



Nous évitons également de collecter des données excessives ou superflues qui ne sont pas directement liées aux finalités spécifiques pour lesquelles les données sont collectées.

Nous supprimons les 13 variables concernées :

Métier	Formation
Nom employeur	Est_rouge
Numéro de sécurité sociale	Date_demande
Id_site_web	Etat_dossier
Nom	Lat
E-Mail	lon
Nombre enfants	

Nous procédons ensuite à l'anonymisation des données.

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible.

Les techniques d'anonymisation à mettre en place peuvent être regroupées en deux familles : la randomisation et la généralisation.

La randomisation consiste à modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale.

Cette technique permet de protéger le jeu de données du risque d'inférence qui est une méthode d'analyse ou de traitement des données qui permet de déduire des informations supplémentaires ou sensibles sur des individus à partir des données collectées, même si ces informations ne sont pas directement révélées ou incluses dans les données elles-mêmes.

La généralisation consiste à modifier l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes.

Cette technique permet d'éviter l'individualisation d'un jeu de données.

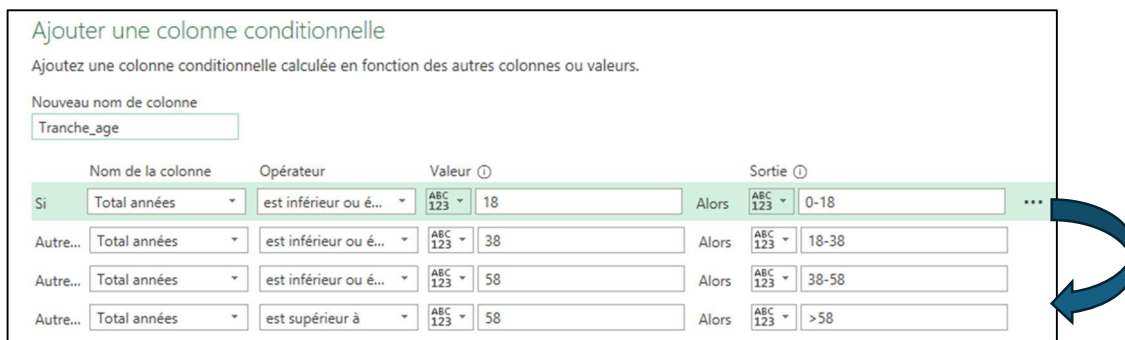
Elle limite également les possibles corrélations du jeu de données avec d'autres.

## **Anonymisation de l'âge :**

Nous calculons l'âge par rapport à la date de naissance par la fonction Age.

Power Query nous donne le nombre de jours qu'on convertit en nombre d'années.

Ensuite, on crée une colonne Tranche d'âge avec le Menu «Ajouter une colonne / colonne conditionnelle» et en insérant les formules pour constituer les tranches d'âge.



**Ajouter une colonne conditionnelle**  
Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne  
Tranche\_age

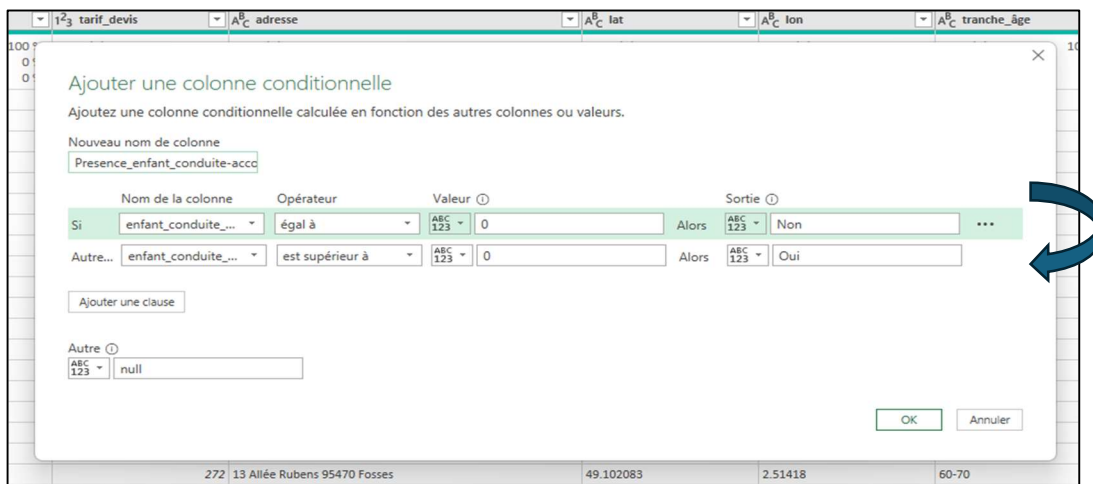
	Nom de la colonne	Opérateur	Valeur ①		Sortie ②
Si	Total années	est inférieur ou é...	ABC 123 18	Alors	ABC 123 0-18
Autre...	Total années	est inférieur ou é...	ABC 123 38	Alors	ABC 123 18-38
Autre...	Total années	est inférieur ou é...	ABC 123 58	Alors	ABC 123 38-58
Autre...	Total années	est supérieur à	ABC 123 58	Alors	ABC 123 >58

On supprime ensuite les 3 colonnes Date de naissance, Age et Total années.

## **Anonymisation du nombre d'enfant en conduite accompagné :**

Nous remplaçons le nombre d'enfant en conduite accompagné par une valeur binaire pour indiquer leur présence ou non (Oui/Non).

Puis nous supprimons la colonne enfant\_conduite\_accompagne.



**Ajouter une colonne conditionnelle**  
Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne  
Presence\_enfant\_conduite-acc

	Nom de la colonne	Opérateur	Valeur ①		Sortie ②
Si	enfant_conduite_...	égal à	ABC 123 0	Alors	ABC 123 Non
Autre...	enfant_conduite_...	est supérieur à	ABC 123 0	Alors	ABC 123 Oui

Ajouter une clause

Autre ①  
ABC 123 null

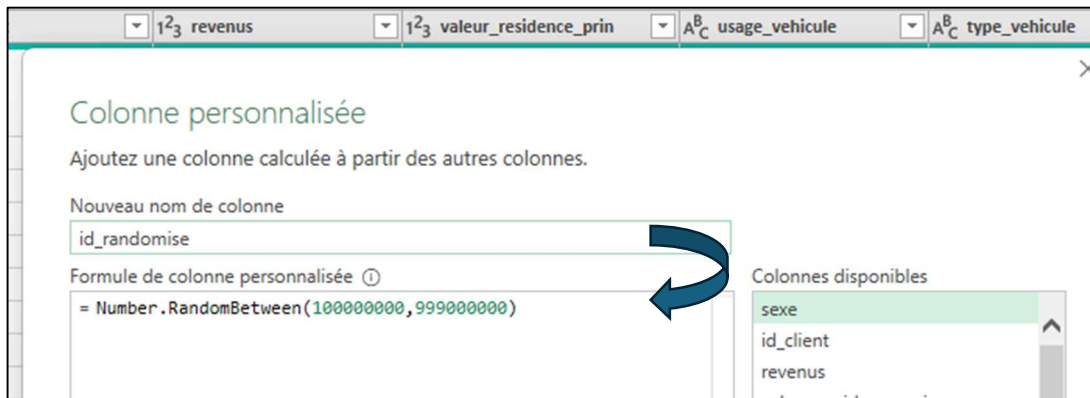
OK Annuler

## Anonymisation de l'identifiant client :

Nous remplaçons l'identifiant client par un index aléatoire.

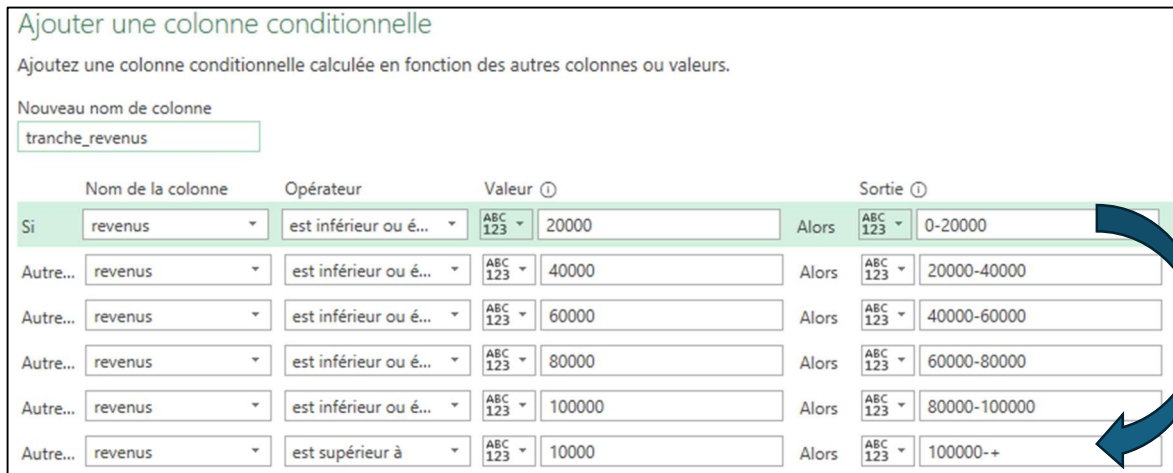
Menu «Ajouter une colonne / Colonne personnalisée» et nous utilisons la fonction «Number.RandomBetween(100000000,999000000) ».

Puis nous supprimons la colonne id\_client.



## Anonymisation des revenus :

Avec la fonction Colonne conditionnelle, nous insérons la formule pour remplacer les revenus par un intervalle ([0, 20k€], [20k€, 40k€], [40k€,60k€], [60k€, 80k€], [80k€, 100k€], [100k€et plus].



	Nom de la colonne	Opérateur	Valeur ①		Sortie ①
Si	revenus	est inférieur ou é...	20000	Alors	0-20000
Autre...	revenus	est inférieur ou é...	40000	Alors	20000-40000
Autre...	revenus	est inférieur ou é...	60000	Alors	40000-60000
Autre...	revenus	est inférieur ou é...	80000	Alors	60000-80000
Autre...	revenus	est inférieur ou é...	100000	Alors	80000-100000
Autre...	revenus	est supérieur à	10000	Alors	100000-+

Nous supprimons ensuite la colonne revenus.





### **Anonymisation de la valeur résidence principale :**

A partir de cette colonne et toujours avec la colonne conditionnelle, nous déduisons si le client est locataire ou propriétaire.

Si la valeur de la résidence principale est de 0, le client est locataire.

Choix binaire, si la valeur est supérieure à 0, on estimera que le client est propriétaire.

Nous supprimons ensuite la colonne valeur\_resid\_principale.

### **Anonymisation des points perdus**

Avec la fonction colonne conditionnelle, si le client a la totalité des points, le champ de la cellule sera rempli par Totalité.

Au cas contraire, le champ sera rempli par Points\_perdus.

Nous conservons uniquement l'information si le client a des points perdus ou non.

Nous supprimerons la colonne points\_perdus.

### **Anonymisation des tarifs devis :**

Pour empêcher que l'équipe de performance commerciale n'ait accès aux données tarifaires détaillées appliquées par les équipes, nous regrouperons les tarifs par tranche : [0€-250€], [250€-500€], [500€-plus].

Avant de rajouter la colonne conditionnelle, il faut remplacer les points par des virgules.

**Ajouter une colonne conditionnelle**

Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne  
Tranche\_tarif\_devis

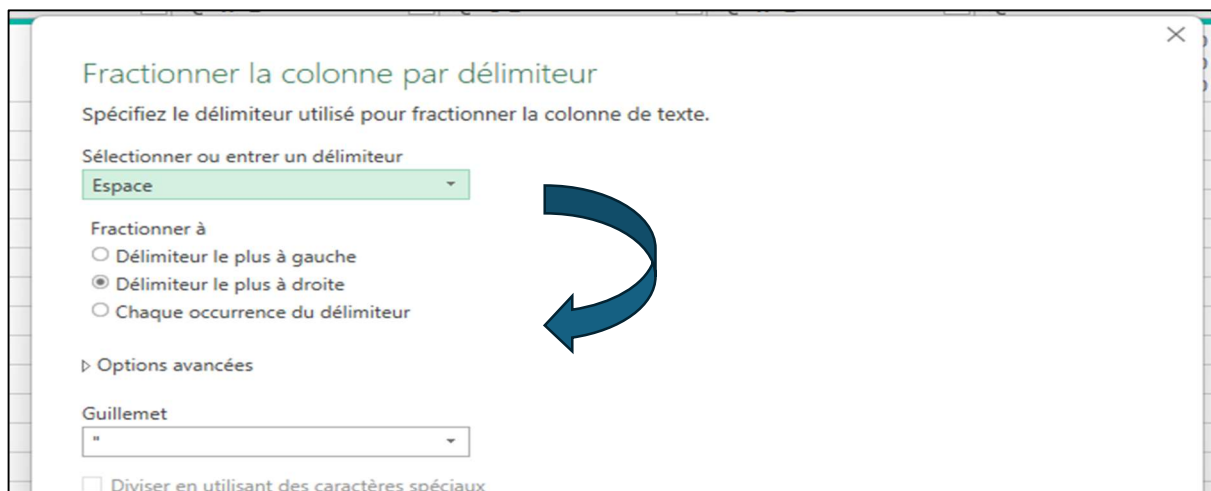
	Nom de la colonne	Opérateur	Valeur ①		Sortie ①
Si	tarif_devis	est inférieur ou é...	ABC 123 250	Alors	ABC 123 0-250
Autre...	tarif_devis	est inférieur ou é...	ABC 123 500	Alors	ABC 123 250-500
Autre...	tarif_devis	est supérieur à	ABC 123 500	Alors	ABC 123 500-+

Nous supprimons ensuite la colonne des tarifs détaillés.

## **Anonymisation des adresses :**

Nous ne conservons que les noms des villes dans le fichier en utilisant la fonction Fractionner la colonne avec un espace comme délimiteur et en cochant le délimiteur le plus à droite.

Nous renommons cette colonne nom\_ville et nous supprimons la colonne Adresse



Ce sera la dernière étape du traitement des données personnelles.

Les variables qui n'ont pas été modifiées ne peuvent pas permettre l'identification des clients : sexe, usage véhicule, type véhicule, âge véhicule, type conduite, formule assurance.

Elles n'ont pas été supprimées car elles peuvent être nécessaires pour l'analyse commerciale de la clientèle.

## **Conclusion**

*Dev'Immédiat doit se mettre rapidement en conformité par rapport au RGPD sinon sa pérennité risque d'être mise en cause. D'une part, parce que l'entreprise ne pourra pas piloter son activité avec la suspension de son fichier CRM et d'autre part, elle n'est pas à l'abri d'un risque d'une lourde sanction financière pour non-respect de la protection des données personnelles.*

*Les différentes étapes décrites dans le rapport répondent aux objectifs de mise en conformité du traitement des données personnelles, mais elles ne sauraient être suffisantes car il faut une démarche continue en faisant évoluer les procédures et les mesures de sécurité en permanence par rapport à une réglementation et un environnement qui évoluent également.*