



# Rapport de traitement des données

<b>Titre du projet</b>	<i>Collectez des données en respectant les normes RGPD</i>
------------------------	--

Version	Auteur	Description	Date
V1	<i>P.Boireau</i>	<i>Rapport</i>	<i>01.04.2023</i>

Introduction
<i>Suite à une plainte d'un client, l'entreprise a été sanctionnée par la CNIL d'une limitation temporaire des traitements. Les employés n'ont plus le droit de manipuler des données personnelles pendant 6 mois. Pour que les équipes puissent continuer à travailler, il faut élaborer une procédure d'extraction des données clients, adaptée.</i>



### Contenu du rapport :

Voici les principales étapes à effectuer pour anonymiser les données de la base Client et obtenir des données avec lesquelles vous pourrez travailler pendant la période de sanction de la CNIL.

- **Etapes pour extraire les données de la base de Données:**

o **Requête SQL** de la Base de Données :

```
SELECT  
    strftime('%Y', date_demande) as annee_demande,  
    sexe, enfant_conduite_accompagne, nombre_enfants,  
    usage_vehicule, type_vehicule, est_rouge, revenus,  
    points_perdus, age_vehicule, type_conduite, etat_dossier,  
    formule, id_client, tarif_devis
```

```
FROM base_client
```

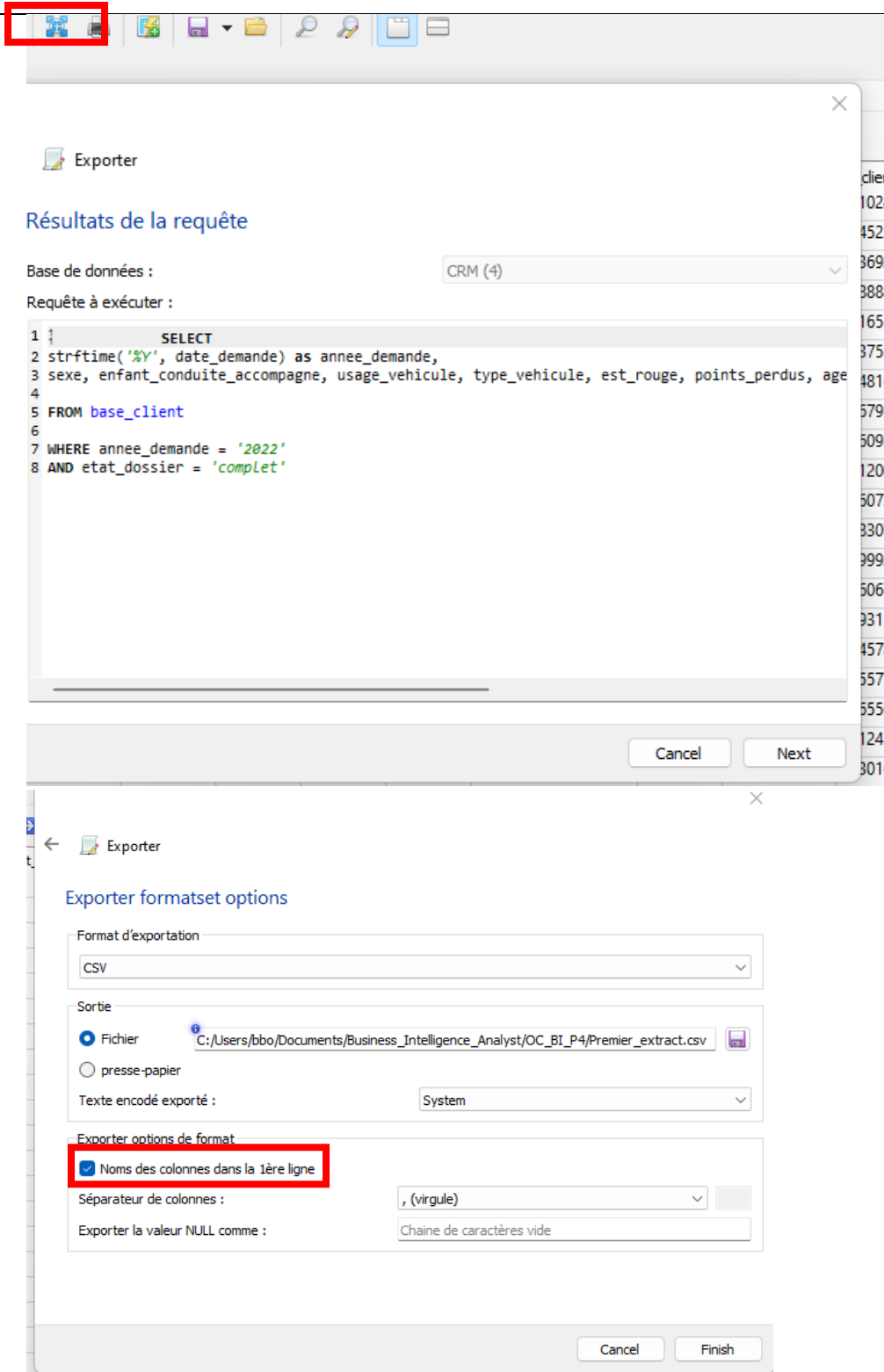
```
WHERE annee_demande = '2022'
```

```
AND etat_dossier = 'complet'
```

▪ Cette requête permet d'obtenir des données anonymisées des dossiers (sans les données permettant d'identifier directement les personnes), filtrées sur les dossiers complets de l'année 2022

▪ Il était aussi possible d'extraire uniquement l'année de naissance des clients en complétant cette requête, mais la colonne de données obtenue apparaît fausse (exemple: plusieurs clients auraient moins de 10 ans, un client né en 1931 aurait un enfant en conduite accompagnée dans son contrat...); il faudra retravailler ces données avec les responsables de leur collecte

o **Exporter** les résultats de la requête SQL dans un fichier csv:



**Exporter**

**Résultats de la requête**

Base de données : CRM (4)

Requête à exécuter :

```

1 SELECT
2 strftime('%Y', date_demande) as annee_demande,
3 sexe, enfant_conduite_accompagne, usage_vehicule, type_vehicule, est_rouge, points_perdus, age
4
5 FROM base_client
6
7 WHERE annee_demande = '2022'
8 AND etat_dossier = 'complet'

```

Cancel Next

**Exporter formatset options**

Format d'exportation: CSV

Sortie:

☒ Fichier C:/Users/bbo/Documents/Business\_Intelligence\_Analyst/OC\_BI\_P4/Premier\_extract.csv

☐ presse-papier

Texte encodé exporté : System

**Exporter options de format**

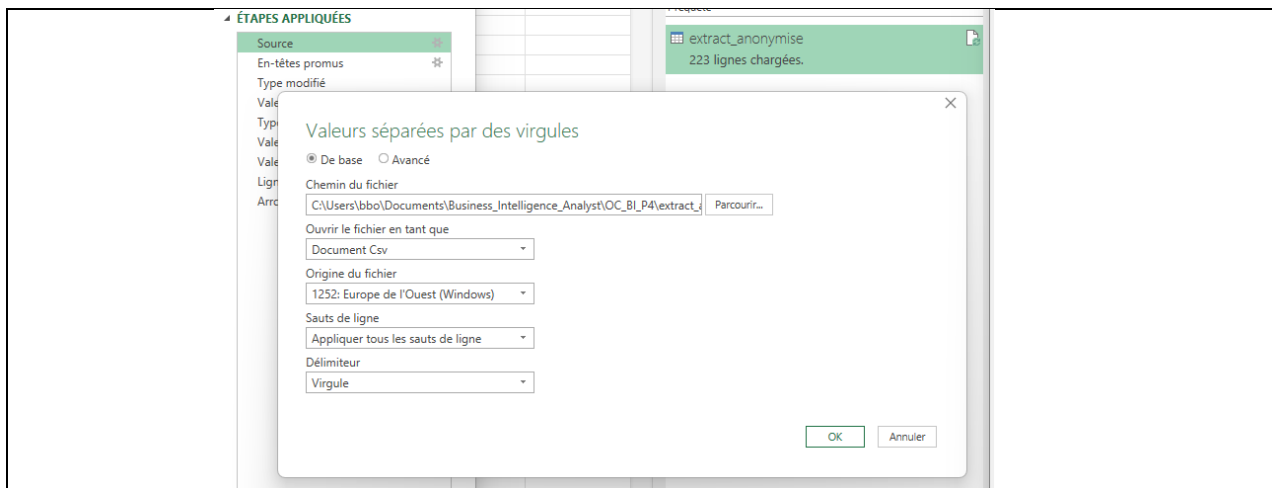
☒ Noms des colonnes dans la 1ère ligne

Séparateur de colonnes : , (virgule)

Exporter la valeur NULL comme : Chaîne de caractères vide

Cancel Finish

- **Importer** les données du fichier csv dans un fichier **excel**:  
Dans Excel, en haut, "Données"/"A partir d'un fichier texte csv", remplir la fenêtre qui s'affiche comme suit:



- **Transformer les données brutes** dans le fichier Power Query:
  - A droite dans Requêtes et Connexions, se positionner sur l'icône du fichier csv lié, clique droit/modifier les données;
    - Le fichier Power Query s'affiche
  - Travailler les données avec Power Query:
    - Se positionner sur la colonne 'tarif\_devis'; clique droit/remplacer les valeurs; remplacer les points par des virgules:

= Table.TransformColumnTypes("#Valeur remplacée",{{"tarif\_devis", type number}})

Index	sexe	enfant_conduite_accompagne	usage_vehicule	type_vehicule
1	F	0	Commercial	Sports Car
2	F	1	Private	Minivan
3	F	0	Private	Pickup
4	F	0	Private	z_SUV
5	M	0	Commercial	Panel Truck
6	F	0	Private	z_SUV
7	F	2	Commercial	Minivan
8	M	0	Private	z_SUV
9	M	0	Private	Minivan
10	F	0	Commercial	Panel Truck
11	F	1	Private	Minivan

Paramètres d'une requête

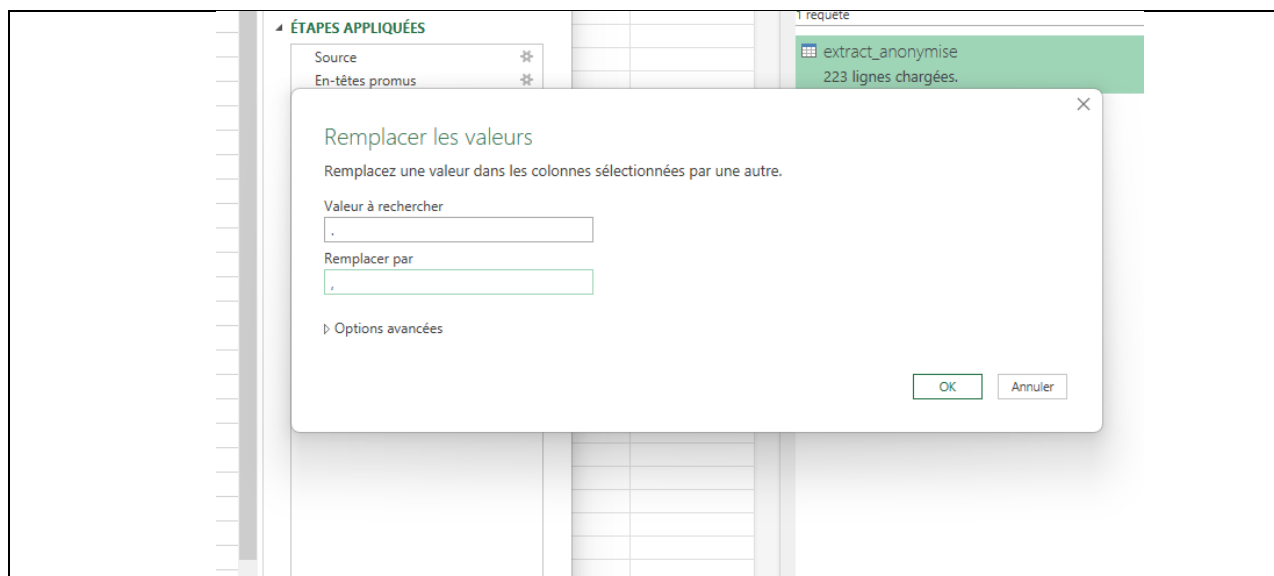
**PROPRIÉTÉS**

Nom: extract\_anonymise

Toutes les propriétés

**ÉTAPES APPLIQUÉES**

- Source
- En-têtes promus
- Valeur remplacée
- Type modifié1**
- Lignes groupées
- Arrondi à l'entier supérieur



- Puis: clique droit/modifier le type; 'Nombre Décimal'

```
= Table.TransformColumnTypes("#Valeur remplacée",{{"tarif_devis", type number}})
```

	A <sup>B</sup> <sub>C</sub> etat_dossier	A <sup>B</sup> <sub>C</sub> formule	A <sup>B</sup> <sub>C</sub> id_client	1.2 tarif_devis
an	complet	dev_integral	871024631	323,39
an	complet	dev_express	7945239	257,56
ral	complet	dev_express	713692930	250,68
an	complet	dev_integral	588888550	300,3
an	complet	dev_vip	69165559	648,13
ral	complet	dev_premium	573751437	459,83
ral	complet	dev_express	294816298	261,28
an	complet	dev_premium	686791659	343,17
an	complet	dev_premium	866093378	389,24
an	complet	dev_integral	401200262	310,66
an	complet	dev_express	926073051	261,6

Paramètres d'une requête...

#### PROPRIÉTÉS

Nom

extract\_anonymise

Toutes les propriétés

#### ÉTAPES APPLIQUÉES

Source

En-têtes promus

Valeur remplacée

✕ Type modifié1

Lignes groupées

Arrondi à l'entier supérieur

- Puis: en haut, „Ajouter une colonne“/“Colonne conditionnelle et remplir les conditions comme suit, pour **transformer** les données de **tarif\_devis** en intervalles de tarifs de 100 entre 200 et 1000:

## Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ②
Si	tarif_devis	est inférieur à	ABC 123 300	Alors	ABC 123 ""200-300""
Autre...	tarif_devis	est inférieur à	ABC 123 400	Alors	ABC 123 ""300-400""
Autre...	tarif_devis	est inférieur à	ABC 123 500	Alors	ABC 123 ""400-500""
Autre...	tarif_devis	est inférieur à	ABC 123 600	Alors	ABC 123 ""500-600""
Autre...	tarif_devis	est inférieur à	ABC 123 700	Alors	ABC 123 ""600-700""
Autre...	tarif_devis	est inférieur à	ABC 123 800	Alors	ABC 123 ""700-800""

Ajouter une clause

Autre ①





- Pour avoir une colonne conditionnelle pour grouper **l'âge des véhicules** par catégorie, faire pareil que précédemment en remplissant les conditions suivantes:

## Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ②
Si	age_vehicule	égal à	ABC 123 null	Alors	ABC 123 ""non renseigné""
Autre...	age_vehicule	est inférieur à	ABC 123 5	Alors	ABC 123 ""0-5ans""
Autre...	age_vehicule	est inférieur à	ABC 123 10	Alors	ABC 123 ""5-10""
Autre...	age_vehicule	est supérieur à	ABC 123 9	Alors	ABC 123 ""10 ans et plus""

Ajouter une clause

Autre ①

- Pour avoir une colonne conditionnelle pour grouper les **Revenus** par catégorie, faire pareil que précédemment en remplissant les conditions suivantes:

Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ①	
Si	revenus	égal à	ABC 123 null	Alors	ABC 123 ""non renseigné""	...
Autre...	revenus	est inférieur à	ABC 123 10000	Alors	ABC 123 ""Moins de 10 K""	
Autre...	revenus	est inférieur à	ABC 123 50000	Alors	ABC 123 ""50k - 100k""	
Autre...	revenus	est inférieur à	ABC 123 200000	Alors	ABC 123 ""100k - 200k""	
Autre...	revenus	est supérieur à	ABC 123 200000	Alors	ABC 123 ""plus de 200k""	

Ajouter une clause

Autre ①  
ABC 123 null

OK Annuler

- Pour avoir une colonne conditionnelle sur la présence **d'enfant**:

Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ①	
Si	nombre_enfants	est supérieur à	ABC 123 0	Alors	ABC 123 ""Oui""	...
Autre...	nombre_enfants	est inférieur à	ABC 123 1	Alors	ABC 123 ""Non""	

Ajouter une clause

Autre ①  
ABC 123 null

OK Annuler

- Pour avoir une colonne conditionnelle pour savoir si présence de **conduite accompagnée**:



### Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

	Nom de la colonne	Opérateur	Valeur ①		Sortie ②	
Si	<input type="text" value="enfant_conduite_..."/>	<input type="text" value="égal à"/>	<input type="text" value="0"/>	Alors	<input type="text" value="Non"/>	...
Autre...	<input type="text" value="enfant_conduite_..."/>	<input type="text" value="est supérieur à"/>	<input type="text" value="0"/>	Alors	<input type="text" value="Oui"/>	

Autre ①

- Puis: retourner sur le bandeau haut dans "Transformer"/ "Regrouper par";  
Pour avoir des données pratiques à manipuler et totalement anonymes, choisir les **regroupements** comme dans les impr écran suivants (vous pouvez choisir de regrouper autant de colonnes que vous le souhaitez; attention plus vous choisissez de colonne, moins vos données seront anonymisées, car l'information devient plus précise, et concerne moins d'individus):



## Regrouper par

Spécifiez les colonnes de regroupement et une ou plusieurs sorties.

☐ De base ☒ Avancé

sexe

usage\_vehicule

type\_vehicule

type\_conduite

formule

enfant\_conduite\_accompagne

Ajouter un regroupement

Nouveau nom de colonne

Nombre

Ajouter une agrégation

Opération

Compter les lignes

Colonne

OK

×

## Regrouper par

Spécifiez les colonnes de regroupement et une ou plusieurs sorties.

☐ De base
 ☒ Avancé

type\_vehicule ▾

type\_conduite ▾

formule ▾

enfant\_conduite\_accompagne

est\_rouge ▾

tarif\_intervalle ▾

Ajouter un regroupement

Nouveau nom de colonne

Opération

Colonne

Nombre

Compter les lignes ▾

Ajouter une agrégation

OK Annuler

## Conclusion

*Afin de préserver au mieux les données personnelles des clients, il convient:*

- *De s'assurer de quelles données l'entreprise a absolument besoin pour fonctionner, et de ne pas collecter plus de données personnelles que nécessaires*
- *De s'assurer que les données personnelles d'un individu sont stockées de manière fractionnée, cryptées pour certaines, et dans des endroits très sécurisés, pour ne pas exposer les clients au piratage de leurs données*
- *D'avoir une gestion très stricte des permissions d'accès aux données personnelles des clients dans l'entreprise, les collaborateurs ne devant avoir accès qu'aux données personnelles dont ils ont absolument besoin*



- *Dès que possible, essayer de “flouter” les données personnelles (par des regroupements, des intervalles,...)*

*Le traitement des données personnelles est un enjeu de sécurité nationale. En effet des entités mal intentionnées peuvent utiliser les données personnelles pour s'introduire par effraction dans les systèmes des entreprises ou des administrations, pour faire pression sur des individus en situation de prise de décision, ...*

*La cybersécurité du pays est l'affaire de tous les citoyens.*