

## RAPPORT DE TRAITEMENT DES DONNÉES

### PROJET : COLLECTER DES DONNÉES EN RESPECTANT LES NORMES RGPD

**Auteur : Sofiane AYOUB**

**Date : 16/11/2023**

### 1/ CONTEXTE DU PROJET

La mission consiste à aider **Dev'immédiat**, une filiale d'un groupe d'assurance, à traiter les dossiers de demande d'assurance automobile reçus et à réaliser la **collecte des données** en respectant le **Règlement Général de Protection des Données**.

Suite à la **plainte d'un client**, la filiale a été sanctionnée par la **CNIL d'une limitation temporaire des traitements**. Mon rôle en tant que consultant analyste BI consiste à **s'assurer que les règles du RGPD sont parfaitement appliquées et intégrées** dans les processus pour demander la levée de la **sanction**.

Le **plan d'action** mis en place pour mener ce projet est le suivant :

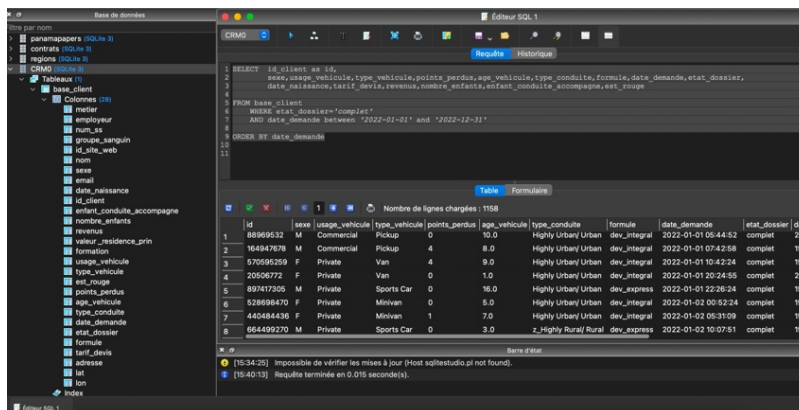
- Première phase, émettre des **recommandations sur les règles de gestion** des données du CRM afin de **respecter** la conformité au **RGPD**.
- Deuxième phase, effectuer une **extraction des données** du CRM au format CSV en les **anonymisant** avec l'outil **Power Query** afin d'éviter un contrôle de la CNIL.
- Dernière phase, **rédiger un rapport** présentant les **étapes des opérations** et démontrant le respect des processus qualité dans la **collecte**, la **préparation des données** et les **préconisations**.

### 2/ PRÉCONISATIONS LIÉES AU RGPD

- **Sélectionner** les champs strictement **utiles et nécessaires** en les « **anonymisant** » ;
- **Combiner** différentes **techniques d'anonymisation** du jeu de données (par exemple remplacer le champ « **id\_client** » par une **table d'index**, utiliser des **opérateurs booléens OUI/NON** ou **VRAI/FAUX**, **randomiser** les attributs en les permutant, **généraliser** le jeu de données en jouant sur les **ordres de grandeurs** afin d'éviter l'individualisation, par exemple en créant des **tranches de revenus [<20k], [20k-40k]** ;
- **Respecter le droit des personnes** concernant la maîtrise de leurs données (**droit d'accès**, de **rectification**, de **suppression**, d'**opposition**, **droit à la limitation du traitement**).

### 3/ DOCUMENTATION SUR LE TRAITEMENT DES DONNÉES

- **Requête SQL des dossiers complet pour l'année 2022**



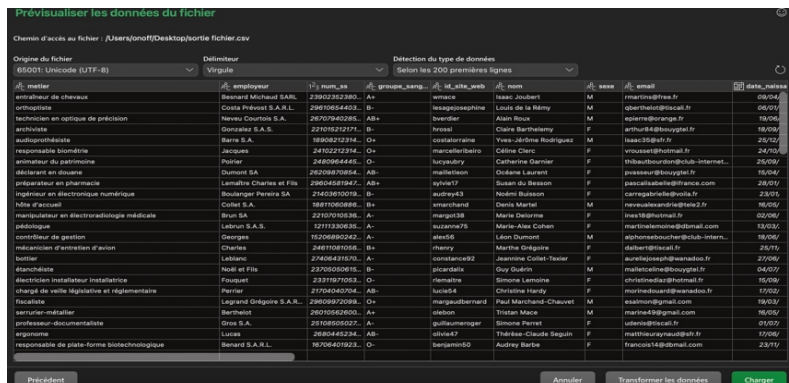
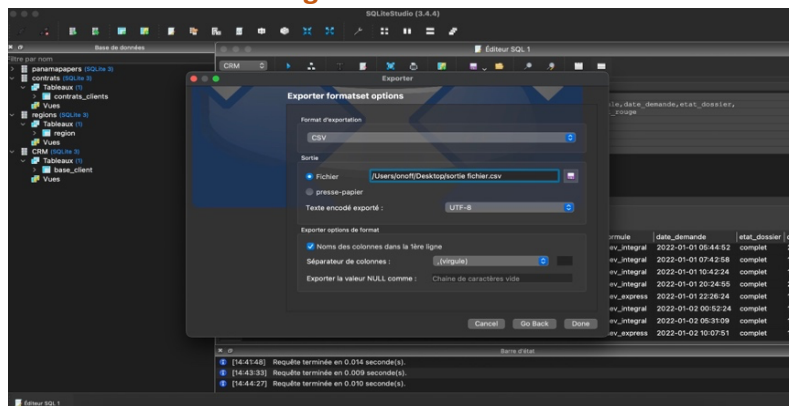
The screenshot shows a SQL editor with a query and its results. The query is as follows:

```
SELECT id_client as id,
       id_voiture, type_voiture, points_perdus, age_voiture, type_conduite, formule, data_demande, etat_dossier,
       data_naissance, tarif_devls, revenus, nombre_enfants, enfant_conduite, accompagnateur, est_sous
FROM base_client
WHERE etat_dossier='complet'
AND data_demande BETWEEN '2022-01-01' AND '2022-12-31'
ORDER BY data_demande
```

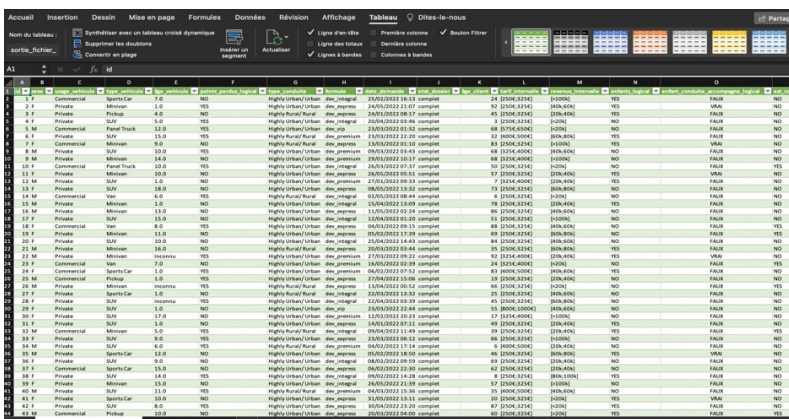
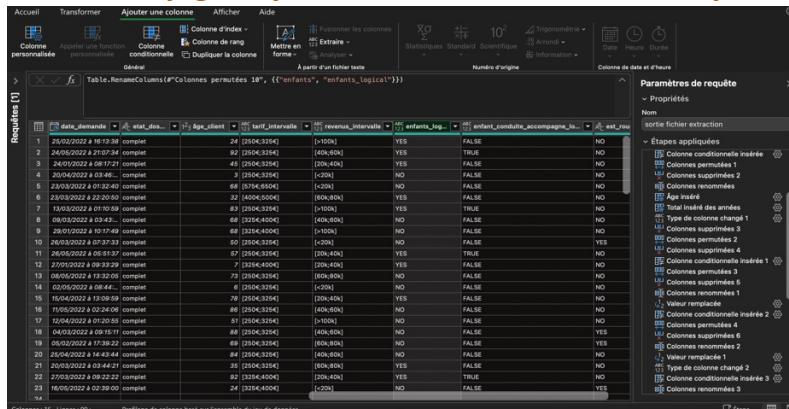
The results table shows 8 rows of data. The columns are: id, sexe, usage\_voiture, type\_voiture, points\_perdus, age\_voiture, type\_conduite, formule, data\_demande, etat\_dossier, and date.

id	sexe	usage_voiture	type_voiture	points_perdus	age_voiture	type_conduite	formule	data_demande	etat_dossier	date
1	88969532	M	Commercial	Pickup	0	10.0	Highly Urban/Urban	dev_integral	2022-01-01 06:44:52	complet
2	164947678	M	Commercial	Pickup	4	8.0	Highly Urban/Urban	dev_integral	2022-01-01 07:42:58	complet
3	570595259	F	Private	Van	4	9.0	Highly Urban/Urban	dev_integral	2022-01-01 10:42:24	complet
4	20506772	F	Private	Van	0	1.0	Highly Urban/Urban	dev_integral	2022-01-01 20:24:55	complet
5	897477325	M	Private	Sports Car	0	16.0	Highly Urban/Urban	dev_express	2022-01-01 22:26:24	complet
6	523698470	F	Private	Minivan	0	5.0	Highly Urban/Urban	dev_integral	2022-01-02 08:52:24	complet
7	44584436	F	Private	Minivan	1	7.0	Highly Urban/Urban	dev_integral	2022-01-02 05:31:09	complet
8	664499270	M	Private	Sports Car	0	3.0	z_Highly Rural/Rural	dev_express	2022-01-02 10:07:51	complet

## • Export au format csv et chargement des données brutes vers Power Query



## • Nettoyage du jeu de données avec Power Query



## 3/ CONCLUSION

En résumé le **RGPD**, mis en application depuis le 25 mai 2018, permet de mieux encadrer le traitement des données personnelles dans l'Union européenne, notamment grâce à l'utilisation de **techniques d'anonymisation** des bases de données (généralisation, randomisation). Celles-ci sont plus ou moins fiables selon les **méthodes** mises en œuvre (ajout de bruit, permutation, confidentialité différentielle, k-anonymisation, pseudonymisation etc). C'est la raison pour laquelle il est recommandé de combiner plusieurs d'entre elles lorsque l'on souhaite anonymiser des données personnelles.