



Rapport du traitement des données

Titre du projet	<i>Extraction & anonymisation des données client</i>
------------------------	--

Version	Auteur	Description	Date
V1	AD	<i>Rapport sur le traitement des données du client</i>	13.05.2024

Introduction

L'entreprise Dev'Immediat, courtier en assurance automobile a été sanctionnée par la CNIL pour ne pas avoir respecté le RGPD.

Elle craint une lourde amende et a besoin de se mettre en conformité dans les plus brefs délais afin de lever la sanction.

Par ailleurs, elle doit pouvoir continuer de piloter son activité commerciale malgré la sanction qui impacte son usage des données.

Pour ce faire, il est attendu de :

- fournir cinq recommandations afin de se conformer au RGPD
- fournir une extraction des données du CRM anonymisées et avec un niveau de détail des données tarifaires réduit (par tranche de tarifs)
- documenter les étapes, objet de ce rapport.

La demande a été formulée par la directrice, Mme Clara Daucour. Mr Jean Luc, son adjoint, sera le DPO (Data Protection Officer) dans le cadre de cette mission.



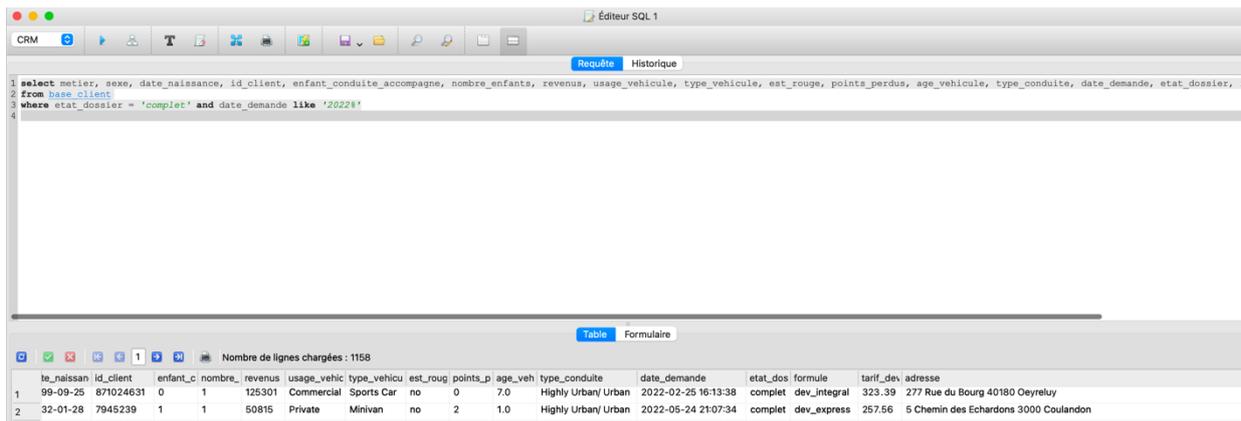
Contenu du rapport

I. Préconisations en lien avec les règles RGPD permettant de garantir à l'avenir le respect du RGPD

- **ne pas collecter les données de santé des clients.** Sauf cas d'usages spécifiques, le traitement de ces données est interdit (art. 9 du RGPD).
- **ne pas collecter les données qui ne sont pas strictement nécessaires** à atteindre la finalité du traitement des données (art. 5 du RGPD).
- **limiter la durée de conservation des données.**
Il s'agit d'un principe général énoncé à l'article 5 du RGPD qui impose à chaque responsable de traitement de déterminer une durée de conservation des données personnelles cohérente et justifiée au regard de l'objectif de leur traitement.
- **documenter la conformité de l'entreprise en matière de conservation de données.**
La documentation constitue un élément clé de conformité aux règles relatives à la protection des données personnelles : les actions entreprises, les analyses menées et les documents réalisés à chaque étape doivent être regroupés et centralisés pour la prouver.
- **sécuriser les données et leur traitement** (art 32 du RGPD).
S'assurer de la mise en place des mesures de sécurité techniques nécessaires à préserver l'intégrité, la disponibilité et la confidentialité des données.
Pour le traitement, compiler dans un document SI de référence « Politique de sécurité des systèmes d'informations (PSSI) », les procédures, contrôles et sensibilisations mises en place pour garantir la sécurité des données.

II. Documentation des traitements effectués

A. documentation de la requête SQL



```
1 select metier, sexe, date_naissance, id_client, enfant_conduite_accompagne, nombre_enfants, revenus, usage_vehicule, type_vehicule, est_rouge, points_perdus, age_vehicule, type_conduite, date_demande, etat_dossier, f
2 from base_client
3 where etat_dossier = 'complet' and date_demande like '2022%'
4
```

	id	he_naissan	id_client	enfant_c	nombre	revenus	usage_vehic	type_vehicu	est_roug	points_p	age_veh	type_conduite	date_demande	etat_dos	formule	tarif_devis	adresse
1		99-09-25	871024631	0	1	125301	Commercial	Sports Car	no	0	7.0	Highly Urban/Urban	2022-02-25 16:13:38	complet	dev_integral	323.39	277 Rue du Bourg 40180 Oeyreluy
2		32-01-28	7945239	1	1	50815	Private	Minivan	no	2	1.0	Highly Urban/Urban	2022-05-24 21:07:34	complet	dev_express	257.56	5 Chemin des Echardons 3000 Coulandon

Capture écran SQLite Studio

```
select metier, sexe, date_naissance, id_client,
enfant_conduite_accompagne, nombre_enfants, revenus,
usage_vehicule, type_vehicule, est_rouge, points_perdus,
age_vehicule, type_conduite, date_demande, etat_dossier, formule,
tarif_devis, adresse
from base_client
where etat_dossier = 'complet' and date_demande like '2022%'
```

Données écartées :

- employeur : inutile
- num_ss : donnée sensible
- groupe_sanguin : donnée médicale
- id_site_web : inutile
- nom : anonymisation par le retrait
- email : anonymisation par le retrait
- valeur_residence_prin : inutile
- formation : inutile
- lat : inutile
- lon : inutile

B. documentation des étapes de traitement du fichier Power Query

1. Import des données au format CSV

Choisir une source de données

Tous Fichier Services en ligne Autre

Rechercher



Classeur Excel

Importez des données à partir d'un classeur Microsoft Excel.



Texte/CSV

Importez des données à partir d'un fichier texte ou CSV.



XML

Importez des données d'un fichier XML.



JSON

Importez des données à partir d'un fichier JSON.



Liste SharePoint Online

Importer des données à partir d'une liste Microsoft SharePoint Online.



OData

Importez des données à partir d'un flux OData.



Table vide

Copiez et collez les données dans la table, ou entrez-les manuellement.



Requête vide

Entrez une requête depuis le début.

Prévisualiser les données du fichier

Chemin d'accès au fichier : /Users/alexandre/SynologyDrive/Drive/Alex/documents/3 - formation en cours/Data Analyst/Open Classroom/Data Analyst - Cours/04 - Collectez des donn...

Origine du fichier: 65001: Unicode (UTF-8) | Délimiteur: Virgule | Détection du type de données: Selon les 200 premières lignes

metier	sexe	date_naissa...	id_cil...	enfant_conduite_accomp...	nombre_enfa...	reven...	usage_vehic...	type_vehic...
animateur du patrimoine	F	25/09/1999	871024631	0	1	125301	Commercial	Sports Car
préparateur en pharmacie	F	28/01/1932	7945239	1	1	50815	Private	Minivan
électricien installateur installatrice	F	15/09/1978	713692930	0	2	21204	Private	Pickup
chargé de veille législative et réglementaire	F	17/02/2021	5888885...	0	0	0	Private	z_SUV
conseiller pénitentiaire d'insertion et de probation	M	14/06/1956	69165559	0	0	null	Commercial	Panel Truck
staffeur-ornemaniste	F	11/04/1992	573751437	0	1	79606	Private	z_SUV
technicien électrotechnicien électrotechnicienne	F	18/03/1941	294816298	2	2	130540	Commercial	Minivan
gestionnaire de patrimoine	M	06/03/1956	686791659	0	0	51278	Private	z_SUV
officier de gendarmerie	M	18/02/1956	8660933...	0	0	115044	Private	Minivan
peintre en bâtiment	F	07/03/1974	401200262	0	0	null	Commercial	Panel Truck
intégrateur web	F	13/02/1967	926073051	1	2	24060	Private	Minivan
coordonnateur d'études cliniques	M	14/05/2017	45830738	0	4	20822	Private	z_SUV

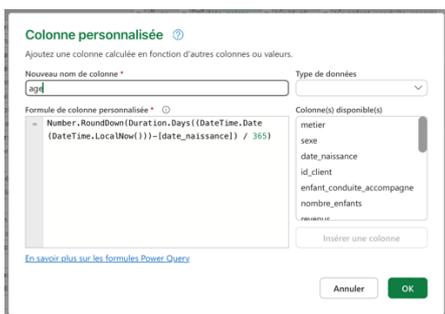
Précédent | Annuler | Transformer les données | Charger

2. Transformation des données

Utilisation de la première ligne de données pour les entêtes.

a) Anonymisation

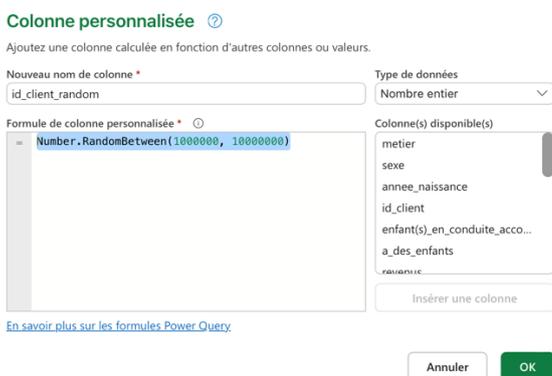
- **date_naissance** remplacée par age



➔ Insertion d'une colonne personnalisée incluant une formule permettant de calculer le nombre de jours séparant la date de naissance du jour actuel, valeur ensuite / 365 et arrondie à l'entier inférieur :

« Number.RoundDown(Duration.Days((DateTime.Date(DateTime.LocalNow()))-[date_naissance]) / 365) »

- **id_client** celui-ci aurait pu être écarté dès l'extraction de la base (ma préconisation). Si toutefois nous le gardons (ce qui est le cas ici)



➔ Création d'une colonne personnalisée nommée « id_client_random » et utilisation de la formule « Number.RandomBetween(1000000, 10000000) » permettant de générer un chiffre aléatoire compris entre les 2 chiffres mis en

parenthèse. Suppression de la colonne d'origine contenant les id.

- **enfant_conduite_accompagnee** nombre remplacé par les valeurs Oui ou Non.

Ajouter une colonne conditionnelle ?

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

Nouveau nom de colonne *

enfant(s)_en_conduite_accom...

Si	Nom de la colonne *	Opérateur *	Valeur ①	Alors	Sortie ①
	1 ² 3 enfant_condui...	est égal à	0		Non
Autrement si	1 ² 3 enfant_condui...	est supérieur à	0		Oui

Ajouter une clause

Else ①

NC

Annuler

OK

➔ Insertion d'une colonne conditionnelle

- **nombre_enfants** remplacé par a_des_enfants
➔ insertion d'une colonne conditionnelle comme ci-dessus
- **revenus** remplacé par tranche_de_revenus ([0-20k€], [20k€-50k€], [50k€-100k€], [100k€-200k€], [+200k€])

Ajouter une colonne conditionnelle ?

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

Nouveau nom de colonne *

tranche_de_revenus

Si	Nom de la colonne *	Opérateur *	Valeur ①	Alors	Sortie ①
	1 ² 3 revenus	est inférieur ou égal à	20000		[0-20k€]
Autrement si	1 ² 3 revenus	est inférieur ou égal à	50000		[20k€-50k€]
Autrement si	1 ² 3 revenus	est inférieur ou égal à	100000		[50k€-100k€]
Autrement si	1 ² 3 revenus	est inférieur ou égal à	200000		[100k€-200k€]
Autrement si	1 ² 3 revenus	est supérieur à	200000		[+ de 200k€]

Ajouter une clause

Else ①

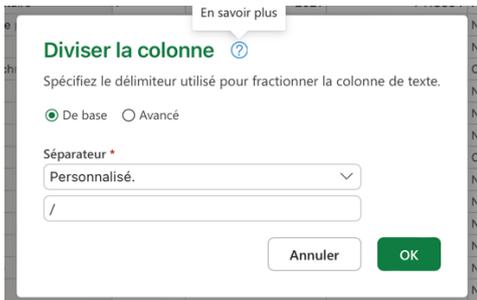
NC

Annuler

OK

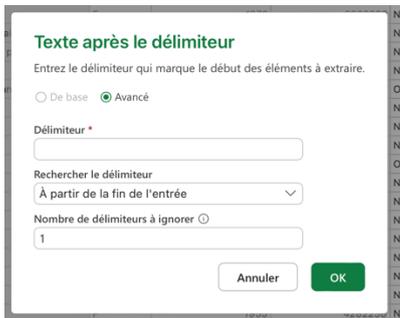
➔ l'insertion d'une colonne conditionnelle

- **date_demande** précise (horodatée) remplacée par mois_demande



→ division de la colonne par délimiteur « / » puis par délimiteur « espace » pour écartier l'heure. Fusion des colonnes mois et année et suppression des colonnes inutiles.

- **adresse** remplacée par localisation (moins précise) ne contenant aucune information sur la rue et le numéro



→ extraction de texte après délimiteur, en partant de la fin et en ignorant un espace pour ne garder que la fin de l'adresse contenant la ville (et éventuellement le CP).

b) Données tarifaires



- **tarif_devis** remplacé par tranche_tarif_devis ([-250€], [250€ à 300€], [300€ à 350€], [350€ à 400€], [400€ à 500€], [+500€],)

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 ②
Si	1 ² ₃ tarif_devis.1	est inférieur à	250	Alors	[- de 250€]
Autrement si	1 ² ₃ tarif_devis.1	est inférieur ou égal à	300	Alors	[de 250 à 300€]
Autrement si	1 ² ₃ tarif_devis.1	est inférieur ou égal à	350	Alors	[de 300 à 350€]
Autrement si	1 ² ₃ tarif_devis.1	est inférieur ou égal à	400	Alors	[de 350 à 400€]
Autrement si	1 ² ₃ tarif_devis.1	est inférieur ou égal à	500	Alors	[de 400 à 500€]
Autrement si	1 ² ₃ tarif_devis.1	est supérieur à	500	Alors	[+ de 500€]

Ajouter une clause

Else ③

NC

Annuler

OK

- ➔ division de la colonne par délimiteur (.) pour arrondir et permettre la reconnaissance en chiffre puis insertion colonne conditionnelle

c) Qualité des données

J'ai profité du traitement des données dans powerQuery pour uniformiser les données en sortie. Ci-dessous le détail des traitements effectués.

- la colonne **est_rouge** utilise des réponse en anglais (yes/no). Je la remplace par une colonne est_rouge_opt avec des réponses en français (oui/non)
 - ➔ insertion de colonne conditionnelle
- la colonne **age_vehicule** comporte un chiffre décimal (.0). Je la simplifie en retirant la décimale. Nouvelle colonne renommée en age_du_vehicule.
 - ➔ division de la colonne avec délimiteur (.) et suppression de la partie inutile



Conclusion

L'ensemble des données transmises a été traité en conformité avec le RGPD.

Les données sensibles et/ou inutiles ont été écartées.

Les données restantes ont été anonymisées afin de permettre leur utilisation tout en respectant la sanction reçue par la CNIL.

Afin d'éviter toute nouvelle sanction pouvant avoir un impact financier lourd pour l'entreprise, il sera important de suivre les recommandations précédemment citées quand au respect du RGPD.

Ces recommandations sont non exhaustives et peuvent évoluer dans le temps.

Il est important pour l'entreprise de se tenir informer (et d'informer/former ses collaborateurs) sur la gouvernance des données et le respect de la réglementation en vigueur.