



Rapport de traitement des données

Titre du projet	<i>Utilisation et protection des données personnelles</i>
------------------------	---

Version	Auteur	Description	Date
V1	Quentin	Rapport	30/06/2022

Introduction
<p>Après discussion avec Clara Daucour et Jean-Luc, les besoins étaient clairs. Il fallait définir un cadre pour le traitement de leurs données personnelles afin d'éviter une nouvelle sanction de la CNIL.</p> <p>Dans un premier temps, il faut émettre les recommandations RGPD à appliquer dans le futur, pour mieux respecter les données personnelles de leurs clients et prospects et ainsi éviter les mécontentements ou les sanctions de la CNIL.</p> <p>Dans un second temps, il faut anonymiser les données pour permettre aux équipes de continuer à piloter la performance commerciale pendant la limitation temporaire de 6 mois.</p>



Contenu du rapport :

- **Préconisations** en lien avec les règles RGPD permettant de garantir à l'avenir le respect du RGPD
 - o Notamment :
 - Consentement client/prospect (opt-in / opt-out)
 - Stocker les données strictement utiles uniquement
 - Suppression auto en cas d'inactivité (3 ans selon la CNIL)
 - Pseudonymisation des données
 - Stratégie data gouvernance pour identifier les rôles avec un data Stewart et un DPO
 - o Voir fichier sur les recommandations RGPD pour avoir plus de détails sur ces recommandations.
- **Documentation** des traitements effectués :
 - o Notamment :
 - Documentation de la requête SQL

Étape 1 : Définir les colonnes à garder et à supprimer.

- Les colonnes supprimées :
 - . Num_ss
 - . Id_site_web
 - . Nom
 - . Email
 - . Lat
 - . Lon
 - . Employeur
 - . Groupe_sanguin
 - . adresse

Ces colonnes ont été supprimées pour l'une de ces 2 raisons :

- L'anonymisation
- L'utilité



Étape 2 : effectuer la requête SQL sur l'année 2022.

```
SELECT  metier,
        sexe,
        date_naissance,
        id_client,
        enfant_conduite_accompagne,
        nombre_enfants,
        revenus,
        valeur_residence_prin,
        usage_vehicule,
        type_vehicule,
        points_perdus,
        age_vehicule,
        type_conduite,
        date_demande,
        etat_dossier,
        formule,
        tarif_devis
FROM    base_client
WHERE   date_demande
BETWEEN "2022-01-01" AND "2022-12-31"
```

Requête

Historique

1

select

metier,

2

sexe,

3

date_naissance,

4

id_client,

5

enfant_conduite_accompagne,

6

nombre_enfants,

7

revenus,

8

valeur_residence_prin,

9

usage_vehicule,

10

type_vehicule,

11

points_perdus,

12

age_vehicule,

13

type_conduite,

14

date_demande,

15

etat_dossier,

16

formule,

17

tarif_devis

18

from

base_client

19

where

date_demande between "2022-01-01" and "2022-12-31"

Table

Formulaire



- Documentation des étapes de traitement du fichier Power Query

Dans un premier temps il a fallu retravailler sur le dictionnaire de données pour modifier les colonnes qui devaient l'être.

Dictionnaire des données modifié :



Nom Champ	Type	Description	Modification effectuée	Nouveau type	Nouvelle description
metier	string	métier du client	INCHANGÉ	string	
employeur	string	employeur du client	SUPPRIMÉ	//	
num_ss	string	numéro de sécurité sociale du conducteur principal	SUPPRIMÉ	//	
groupe_sanguin	string	groupe sanguin du client	SUPPRIMÉ	//	
id_site_web	string	identifiant utilisé pour l'inscription sur notre site internet	SUPPRIMÉ	//	
nom	string	nom et prénom du client	SUPPRIMÉ	//	
sexe	string	sexe	INCHANGÉ	string	
email	string	email de contact du client	SUPPRIMÉ	//	
date_naissance	string	date de naissance	MODIFIÉ	integer	Année de naissance
id_client	string	identifiant unique du client dans notre CRM	MODIFIÉ	integer	identifiant randomisé
enfant_conduite_accompagne	string	nombre d'enfants en conduite accompagnée	INCHANGÉ	boolean	
nombre_enfants	string	nombre d'enfants du client	MODIFIÉ	boolean	présence d'enfants
revenus	string	revenus de référence du client	MODIFIÉ	string	revenus par tranche de 20K
valeur_residence_prin	string	valeur de la résidence principale du client (estimation client)	MODIFIÉ	string	valeur résidence principale par tranche
formation	string	type de formation du client	SUPPRIMÉ	//	
usage_vehicule	string	usage principal du véhicule	INCHANGÉ	string	
type_vehicule	string	type de véhicule	INCHANGÉ	string	
est_rouge	string	indique si le véhicule du client est rouge	SUPPRIMÉ	//	
points_perdus	string	points retirés sur le permis de conduire à la date de la demande d'assurance	INCHANGÉ	string	
age_vehicule	string	âge du véhicule à assurer	INCHANGÉ	string	
type_conduite	string	type de conduite envisagée	INCHANGÉ	string	
date_demande	string	date de la demande de devis faite sur internet	MODIFIÉ	date mois/année	mois et année de la demande
etat_dossier	string	état du dossier : complet / incomplet	INCHANGÉ	string	
formule	string	formule choisie par le client lors de sa demande internet	INCHANGÉ	string	
tarif_devis	string	tarif accordé au client en fonction de son profil	MODIFIÉ	string	tarif par tranche de 100 €
adresse	string	adresse de la résidence principale du client	SUPPRIMÉ	//	
lat	string	latitude de la résidence principale du client	SUPPRIMÉ	//	
lon	string	longitude de la résidence principale du client	SUPPRIMÉ	//	



Modification des colonnes restantes.

- Colonnes modifiées :
 - . Date_naissance ▶ Annee_naissance
 - . Revenus ▶ Tranches_revenus
 - . Valeur_residence_prin ▶ Tranches_valeur_residence_principale
 - . Tarif_devis ▶ Tranche_tarif_devis

Après ces modifications principales, j'ai effectué quelques modifications mineures notamment sur les types de chaque colonne pour être sûr que cela correspond avec les données présentes.

Bien entendu maintenant que l'anonymisation est effectuée, il faut s'entretenir avec les différentes équipes pour déterminer l'utilité, ou non, de chaque donnée dans leur travail afin de respecter la recommandation de data gouvernance.

Si vous souhaitez renforcer encore plus l'anonymisation déjà effectuée sans trop nuire au jeu de données, vous pouvez :

- Récupérer le département correspondant à chaque ville pour n'afficher que le département
- Changer l'intitulé des métiers par leurs codes sociaux professionnels

Cela nécessite d'introduire une nouvelle table dans votre base de données pour y faire correspondre ville/départements et métiers/codes sociaux professionnels.

Bien que cela soit possible, ce n'est cependant pas nécessaire, car le script de Power Query que je vous ai transmis anonymise déjà les données de votre base de données.



Conclusion

Ouverture sur les exigences, contraintes et enjeux de la confidentialité

En respectant les recommandations RGPD dans le document joint, vous n'aurez plus de problèmes avec la CNIL.

La stratégie de Data gouvernance qui sera sous la supervision de Jean-Luc, vous permettra de suivre vos données comme il se doit et cela aura plusieurs avantages :

- *Protection des données*
- *Favorise la défense contre la cybercriminalité*
- *Améliore la relation et la satisfaction client*
- *Gain de temps pour vos équipes*

Pour que cette stratégie soit pérenne, il faudra nommer, en plus d'un data Stewart qui s'occupera d'organiser la base de données pour chaque équipe, un DPO (rôle qui pour l'instant est géré par Jean-Luc) ainsi qu'un intervenant dans chaque équipe afin de se mettre à jour régulièrement sur les besoins business et sur l'évolution de la réglementation.