



Rapport de traitement des données

Titre du projet	Mise en conformité RGPD
-----------------	-------------------------

Version	Auteur	Description	Date
V1	Manon Darget	Rapport de traitement des données	20/02/2024

Introduction

Suite à un signalement de non-conformité RGPD émanant d'un de ses prospects, l'entreprise Dev' Immédiat a reçu une mise en demeure de la CNIL pour non-respect du RGPD avec une sanction de 6 mois sous forme de limitation temporaire des traitements de données. Confrontée à cette contrainte, l'entreprise doit mettre en œuvre des actions correctives et se conformer au RGPD.

Afin de faciliter cette mise en conformité, l'entreprise Dev' Immédiat fait appel à une Entreprise de Services du Numériques (ESN) pour une mission de conseil.

Dans le cadre de cette mission, le directeur adjoint, Jean-Luc assumera, temporairement, le rôle de DPD (Délégué à la Protection des Données).

L'objectif principal de l'entreprise est de réaliser sa **mise en conformité** tout en continuant de **piloter son activité** de prospection commerciale. Pour ce faire, il est primordial de mettre à jour les processus internes relatifs à la collecte, au traitement et à la documentation des données afin de garantir leur exploitation en adéquation avec le RGPD.

Durant la mission de conseil, l'entreprise souhaite obtenir :

Des recommandations sur les règles de gestion à mettre en place pour assurer la conformité des données du CRM.

Une mise à disposition, pour les équipes commerciales, d'une extraction des données du CRM au format CSV, afin de leur permettre de poursuivre leur travail.

Une documentation des étapes d'anonymisation des données (le présent document).



Contenu du rapport :

Préconisations en lien avec les règles RGPD permettant de garantir à l'avenir le respect du RGPD, notamment :

- Collecte (communication d'information vers les clients : transparence, droit des personnes, etc.).
- Traitement des données

Documentation des traitements effectués, notamment :

- Documentation de la requête SQL
- Documentation des étapes de traitement du fichier Power Query
- Documentation des traitements effectués sur la base de données brutes
- Étapes pour retravailler le jeu de données et les explications associées

A. Préconisations

Pour garantir une conformité continue au RGPD, nous vous fournirons un document détaillant les meilleures pratiques à adopter. Ce document exposera en détail les recommandations clés suivantes :

- Désigner un DPD (Délégué à la Protection des Données)
- Constituer un registre de traitements des données
- Minimiser les données
- Respecter les droits des personnes
- Sécuriser ses données

De plus, nous mettons à votre disposition des liens vers des ressources supplémentaires pour approfondir chaque point, dans le but de vous accompagner au mieux dans la mise en place de ces mesures essentielles.

B. Documentation

Afin de fournir aux équipes de performance commerciale une extraction des données du CRM au format CSV, nous avons accédé à la base de données de Dev' Immédiat et exécuté une



requête SQL pour extraire les données de l'année 2022 correspondant aux dossiers qualifiés de complets.

Nous procédons à une extraction via le SGBD (ici SQL Lite) > Exporter > Résultats d'une requête (cf. capture n°1). Une fois la requête saisie, nous sélectionnons une exportation au format CSV dans un fichier que nous nommerons **base_client.csv** (cf. capture n°2).

La requête SQL est la suivante :

```
SELECT * FROM base_client
WHERE etat_dossier = 'complet' AND date_demande LIKE '2022%';
```

← Exporter

Résultats de la requête

Base de données : CRM

Requête à exécuter :

```
1 SELECT * FROM base_client
2 WHERE etat_dossier = 'complet' AND date_demande LIKE '2022%';
```

1 - Exportation de la BDD à l'aide d'une requête SQL

← Exporter

Exporter formatset options

Format d'exportation : CSV

Sortie :

☒ Fichier OC - DATA BI/P4 - Collectez des données en respectant les normes RGPD/base_client.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

2 - Sélection de l'export au format CSV

Avec cette requête nous demandons à sélectionner l'entièreté de la table **base_client** lorsque le paramètre **etat_dossier** contient 'complet' et lorsque **date_demande** commence par 2022. La date étant notée comme suit au sein de notre base de données : AAAA-MM-JJ HH:MM:SS



Après avoir effectué notre extraction, nous avons trié les données en utilisant une méthode de minimisation, conformément à nos recommandations préalables. Pour ce faire, nous avons importé notre fichier CSV au sein d'un classeur Excel que nous avons nommé **base_client.xlsx**. Ce fichier sera notre outil pour les étapes de minimisation et d'anonymisation des données.

Pour commencer, nous avons examiné attentivement les données en nous référant au dictionnaire des données qui nous a été fourni.

Nous avons identifié la présence d'informations sensibles telles que les numéros de sécurité sociale ou les groupes sanguins de vos clients. Conformément à l'article 9 du RGPD, il est interdit de traiter ce type de données sans une base juridique appropriée.

De plus, nous avons constaté que vous collectiez la latitude et la longitude de l'adresse de vos clients. Il est essentiel de noter que la collecte de ces informations n'est autorisée que si vous disposez d'un motif légitime et du consentement explicite de vos clients, comme le stipule l'article 6.1.a et l'article 9 du RGPD.

En outre, nous avons observé que vous demandiez aux clients de préciser si leur véhicule à assurer est de couleur rouge. Cependant, il est important de souligner que l'augmentation des tarifs d'assurance en fonction de la couleur du véhicule pourrait être considérée comme discriminatoire, en violation des réglementations en matière de traitement équitable des clients et de non-discrimination.

Par conséquent, afin de garantir la conformité aux normes légales et éthiques en vigueur, nous avons supprimé les données suivantes :

- Numéro de sécurité sociale
- Groupe sanguin
- Latitude et longitude
- Est rouge

Toujours dans notre démarche de minimisation des données, nous avons procédé à la sélection des informations pertinentes et nécessaires à l'exercice votre activité. Dans ce cadre, nous avons donc supprimé les données suivantes :

- Métier
- Employeur
- Formation
- Valeur Résidence
- Nombre d'enfants
- Id site web
- Adresse e-mail



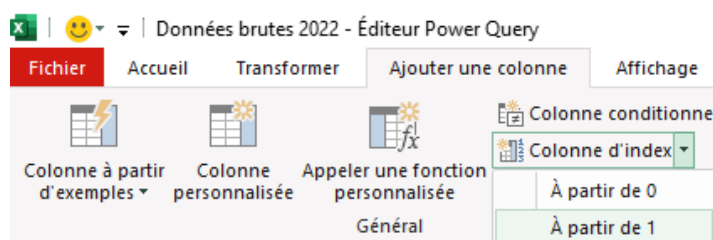
Cette action de minimisation vise à garantir que seules les données essentielles sont conservées.

Nous allons maintenant nous occuper de l'anonymisation des données. Nous avons adopté deux méthodes d'anonymisation : la randomisation des données et la généralisation des données.

La **randomisation des données** consiste à **mélanger** les données de **façon aléatoire** pour protéger la confidentialité tout en préservant leur utilité pour d'éventuelle analyses.

Cette méthode d'anonymisation a été appliquée à deux variables, la variable **nom** et **id_client**

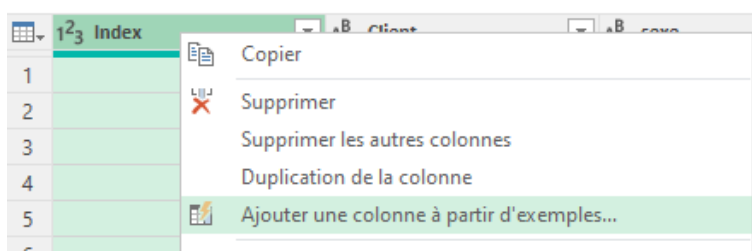
Avec l'outil Power Query, nous avons ajouté une nouvelle colonne d'index pour remplacer l'identifiant client par un index aléatoire. Nous avons ainsi généré des index commençant à partir de 1 pour l'ensemble de notre base de clients (cf. capture n°3).



3 - Ajout d'une colonne d'index

Ensuite, pour anonymiser les noms des clients, nous avons opté pour une dénomination suivant le schéma : 'Client 1', etc. Pour ce faire, nous avons créé une nouvelle colonne indiquant 'Client' suivi de l'index aléatoire déterminé précédemment.

Nous utilisons notre colonne d'index comme base pour créer notre nouvelle colonne. En faisant un clic-droit, nous choisissons *d'ajouter une colonne à partir d'exemple* (cf. capture n°4), ce qui crée une nouvelle colonne. Nous écrivons alors 'Client 1' dans la première ligne (cf. capture n°5) et validons en appuyant sur Entrée, ce qui génère automatiquement nos données (cf. capture n°6).



4 - Ajout d'une colonne à partir de nos index



Colonne1
Client 1

5 - Définition du formalisme de notre colonne

Fusionné
Client 1
Client 2
Client 3
Client 4
Client 5
Client 6
Client 7
Client 8
Client 9
Client 10
Client 11
Client 12
Client 13

6 - Notre colonne auto-générée au formalisme choisi

La **généralisation des données** se résume à **regrouper** les données **pour masquer les détails** afin de préserver la confidentialité tout en permettant l'utilisation pour des fins d'analyses par exemple.

Cette méthode d'anonymisation a été appliquée aux variables : **date_naissance**, **enfant_conduite_accompagne**, **revenus**, **adresse**.

Afin d'anonymiser les dates de naissance, nous avons décidé de ne conserver que l'année de naissance des clients. Pour ce faire, nous avons utilisé la fonctionnalité d'ajout de colonne personnalisée, où nous avons saisi une formule visant à extraire uniquement l'année du champ **date_naissance** (cf. capture n°7).

Colonne personnalisée

Ajoutez une colonne calculée à partir des autres colonnes.

Nouveau nom de colonne

Année de naissance

Formule de colonne personnalisée ①

= Date.Year([date_naissance])

Colonnes disponibles

- Index
- id_client
- nom
- sexe
- email
- date_naissance
- nombre_enfants

<< Insérer

7 - Ajout d'une colonne avec utilisation de la formule Date.Year()

Lors de notre minimisation des données, nous avons supprimé le champ **nombre_enfants** mais avons conservé **enfant_conduite_accompagne**. Afin d'anonymiser les données, nous avons décidé de transformer le nombre d'enfant en une valeur logique (oui ou non). Nous commençons donc par changer le type de données. Actuellement il s'agit d'un nombre entier (le type de données est représenté à côté du nom de notre champ, il suffit de cliquer dessus pour le modifier), nous changeons pour le type Vrai/Faux (cf. capture n°8).

123 enfant_conduite_accompagne ▼

- 1.2 Nombre décimal
- \$ Devise
- 123 Nombre entier
- % Pourcentage
- Date/Heure
- Date
- Heure
- Date/Heure/Fuseau horaire
- Durée
- Texte
- Vrai/Faux

8 - Changement du type de données de enfant_conduite_accompagne

Nous n'avons donc plus des nombres mais la valeur *vrai* si le client dispose d'un enfant en conduite accompagnée. Cependant, pour une meilleure compréhension des équipes, nous avons voulu convertir *Vrai/Faux* en *Oui/Non*. Pour effectuer cette conversion, nous avons créé une colonne conditionnelle à laquelle nous avons attribuer deux clauses (cf. capture n°9) :



- Si au sein de la colonne *enfant_conduite_accompagne* la valeur est égale à *TRUE* (*Vrai*) alors inscrire la mention *Oui*
- Si au sein de la colonne *enfant_conduite_accompagne* la valeur est égale à *FALSE* (*Faux*) alors inscrire la mention *Non*

Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

Enfant conduite accompagnée

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

Ajouter une clause

9 - Ajout d'une colonne conditionnelle pour faciliter la lecture du champ logique

Pour les revenus des clients, nous avons opté pour une classification par catégories de revenus, c'est ce que nous avons fait en ajoutant une colonne conditionnelle. Nous avons effectué le découpage suivant : [€0 - €20K], [€20K - €40K], [€40K - €60K], [€60K - €80K], [€80K - €120K], [€120K +] (cf. capture n°10).

Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

Revenus Annuel

	Nom de la colonne	Opérateur	Valeur ①		Sortie ②	
Si	revenus	égal à	ABC 123 null	Alors	ABC 123 Non renseigné	...
Autre...	revenus	est inférieur ou é...	ABC 123 20000	Alors	ABC 123 €0 - €20K	
Autre...	revenus	est inférieur ou é...	ABC 123 40000	Alors	ABC 123 €20K - €40K	
Autre...	revenus	est inférieur ou é...	ABC 123 60000	Alors	ABC 123 €40K - €60K	
Autre...	revenus	est inférieur ou é...	ABC 123 80000	Alors	ABC 123 €60K - €80K	
Autre...	revenus	est inférieur ou é...	ABC 123 120000	Alors	ABC 123 €80K - €120K	

Ajouter une clause

10 - Ajout d'une colonne conditionnelle pour catégoriser les revenus clients

Concernant le champ *adresse*, nous avons centré notre attention sur la ville de résidence des clients car elle correspond à la ville de stationnement du véhicule. Nous avons donc ajouté une colonne personnalisée et utilisé une formule nous permettant d'extraire la ville des adresses clients (cf. capture n°11).



Colonne personnalisée

Ajoutez une colonne calculée à partir des autres colonnes.

Nouveau nom de colonne
Ville de stationnement

Formule de colonne personnalisée ⓘ
= Text.Middle([adresse], Text.PositionOf([adresse], " ", Occurrence.Last) + 1, Text.Length([adresse]))

Colonnes disponibles
Index
id_client
nom
sexe
email
date_naissance
nombre_enfants
...
<< Insérer

11 - Ajout d'une colonne personnalisé pour extraire la ville de stationnement des véhicules clients

Nous utilisons la formule `Text.Middle()`, qui sert à extraire une partie d'une chaîne de caractères.

Notons que dans notre champ adresse, les villes représentent le dernier élément de notre chaîne de caractère. Les adresses sont notées comme suit : 277 Rue du Bourg 40180 Oeyreluy. Les différents éléments de l'adresse sont séparés d'un espace.

Cette formule prend 3 arguments :

- **Texte à extraire** : chaîne de caractères de laquelle on souhaite extraire une partie. Dans notre cas, `adresse`.
- **Position de départ** : endroit à partir duquel on souhaite commencer l'extraction. Dans notre cas, nous avons utilisé `Text.PositionOf([adresse], " ", Occurrence.Last) + 1` pour trouver l'emplacement du dernier espace de l'adresse, puis nous avons ajouté 1 pour démarrer juste après cet espace.
- **Longueur de la sous-chaîne** : nombre de caractères que l'on souhaite extraire à partir de la position de départ. Dans notre cas, nous avons utilisé `Text.Length([adresse])` pour extraire toute la partie de l'adresse qui se trouve après le dernier espace.

Enfin, pour rendre les données tarifaires moins détaillées et plus accessibles aux équipes de performance commerciale, nous avons de nouveau opté pour une anonymisation par généralisation. Cette fois-ci, nous avons découpé le champ `tarif_devis` en intervalles comme suit : [250 € - 360 €], [360 € - 470 €], [470 € - 580 €], [580 € - 680 €], [680 € - 790 €], [790 € - 890 €], [890 € - 990 €]. Toujours sous le principe d'ajout d'une colonne conditionnelle (cf. capture n°12).



>

Ajouter une colonne conditionnelle

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

Nouveau nom de colonne

Tarification

	Nom de la colonne	Opérateur	Valeur ①		Sortie ①	
Si	tarif_devis	est inférieur ou é...	ABC 123 360	Alors	ABC 123 250 € - 360 €	...
Autre...	tarif_devis	est inférieur ou é...	ABC 123 470	Alors	ABC 123 360 € - 470 €	
Autre...	tarif_devis	est inférieur ou é...	ABC 123 580	Alors	ABC 123 470 € - 580 €	
Autre...	tarif_devis	est inférieur ou é...	ABC 123 680	Alors	ABC 123 580 € - 680 €	
Autre...	tarif_devis	est inférieur ou é...	ABC 123 790	Alors	ABC 123 680 € - 790 €	
Autre...	tarif_devis	est inférieur ou é...	ABC 123 890	Alors	ABC 123 790 € - 890 €	
	ABC		...	

Ajouter une clause

12 - Ajout d'une colonne conditionnelle pour regrouper les tarifs des devis

Conclusion

La démarche de mise en conformité RGPD est une étape cruciale pour l'entreprise Dev'Immédiat, dans un contexte où la confiance des clients et la légitimité des activités commerciales sont en jeu. Malgré les défis et contraintes rencontrés, notamment la nécessité de mettre à jour les processus internes et de sensibiliser les équipes, cette démarche s'avère indispensable. En garantissant la conformité aux normes légales et en adoptant des pratiques de traitement des données respectueuses de la vie privée, l'entreprise renforcera sa crédibilité et sa pérennité sur le marché. La protection des données personnelles est devenue un enjeu majeur dans un environnement numérique en évolution constante, et la démarche de Dev'Immédiat témoigne de sa volonté de répondre à ces défis de manière proactive et responsable.