Modèle de rapport de traitement des données

Titre du projet	Conformisation au RGPD

Version	Auteur	Description	Date
V1	Matthieu Haïdopoulo	Création d'un processus internes sur la collecte, le traitement, l'anonymisation des données clients du CRM pour conformité au RGPD	18/04/2024

Introduction

Suite à une mise en demeure de la CNIL pour non-respect du RGPD (conservation illimitée et non anonymisée des données clients), une sanction a été imposée sous forme de limitation temporaire (6 mois) des traitements, en attendant la mise en place de mesures correctives pour se conformer à la réglementation.
Dans l'immédiat, vous souhaitez pouvoir continuer à piloter votre activité et éviter une sanction de la part de la CNIL. Pour cela, la création d'une équipe dédiée et la nomination d'un DPO ont été décidées afin d'assurer le suivi des recommandations pour se conformer aux règles du RGPD.

DEV'IMMEDIAT

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

Préconisations pour respect du RGPD

Tous d'abord, la décision de former une équipe dédiée est judicieuse. Cette équipe sera en mesure de garantir le suivi des recommandations émises ci-dessous, de fournir des interlocuteurs spécifiques pour toute demande d'information, qu'elle soit interne ou externe, de créer de la valeur grâce à une stratégie data alignée à celle de l'entreprise et de garantir sur le long terme la conformité aux règles RGPD.

• **La minimisation des données :** Uniquement les données essentielles et nécessaires pour atteindre l'objectif fixer à ce traitement peuvent être collecté. Attention il faut porter également une attention particulière sur les données dite sensibles (concernant la santé, les opinions politiques, religieuses, ...) qui ne peuvent être collectées et traitées que dans certaines conditions. Le RGPD interdit de recueillir ou d'utiliser ces données, sauf dans certains cas qui sont précisément listés (cf. Art. 9 du RGPD).

• **La conservation limitée des données :** Il est impératif de déterminer une durée de conservation des données et de l'appliquer rigoureusement. Une fois que l'objectif initial de leur collecte est atteint, les données peuvent être soit archivées, supprimées, ou anonymisées.

DEV'IMMEDIAT

• **La sécurité des données :** Les entreprises sont tenues de mettre en place des mesures de sécurité appropriées pour protéger les données personnelles contre tout accès non autorisé, toute divulgation, toute altération ou toute destruction accidentelle ou illicite via :

-une sécurisation technique : le cryptage, les pare-feux et les antivirus.

-une sécurisation organisationnelle : En plus des mesures techniques, adopter des politiques et des procédures internes pour garantir la sécurité des données, telles que des contrôles d'accès stricts et des formations en matière de sécurité pour le personnel.

-une gestion des risques : mener des évaluations régulières des risques pour identifier les menaces potentielles.

-une notification des violations : En cas de violation de données personnelles, notifier l'autorité de contrôle compétente dans les 72 heures suivant la découverte de la violation, sauf si la violation n'est pas susceptible de présenter un risque pour les droits et libertés des individus concernés.

• **Transparence et droits des personnes :** Les personnes doivent être informées de l'utilisation faites des informations récolter les concernant et la manière dont ils peuvent exercer leurs droits leurs permettant de garder la maitrise de leurs données. Le droit d'accès, le droit de rectification, le droit de suppression, le droit d'opposition, le droit à la portabilité, le droit à la limitation du traitement, le droit de définir le sort des données après la mort, le droit de ne pas faire l'objet d'une décision automatisée.

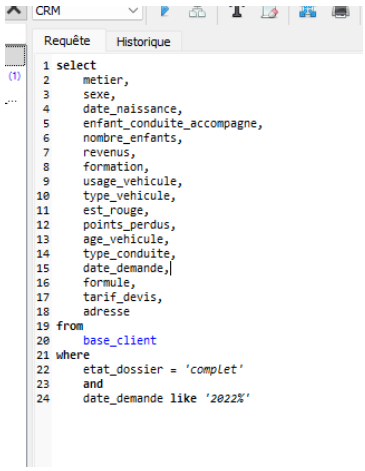
• **La création du registre des activités de traitement :** L'article 30 du RGPD stipule l'obligation pour les responsables du traitement des données de maintenir un registre de leurs activités de traitement. Il doit contenir des informations détaillées sur les activités de traitement des données personnelles effectuées par l'organisation. Cela inclut notamment les finalités du traitement, les catégories de données personnelles traitées, les destinataires des données et les mesures de sécurité mises en place. Le registre est également un outil de responsabilisation pour les responsables du traitement. Il leur incombe de maintenir ce registre à jour et de le tenir à la disposition de l'autorité de contrôle compétente en cas de demande. Il a aussi pour but de contribuer à accroître la transparence des activités de traitement des données au sein de l'organisation. Il permet aux individus de comprendre comment leurs données sont utilisées et traitées. Le registre de l'activité de traitement est un outil essentiel pour les responsables du traitement afin de documenter et de démontrer leur conformité aux obligations en matière de protection des données.

DEV'IMMEDIAT

Documentation des traitements effectués

-Documentation SQL :

Pour commencer, afin de répondre à la problématique de pouvoir continuer à traiter les données malgré la limitation établie par la CNIL, j'ai extrait du CRM, les données de l'année en cours (2022) dont l'état étais complet. Je conservé uniquement les informations nécessaires ainsi que celle que je transformerais par la suite avec Power Query pour anonymisation.
Je n'ai pas gardé les colonnes : employeur, num_ss, groupe_sanguin, id_site_web, nom, email, id_client, valeur_residence_prin, etat_dossier, lat et lon. En effet, ces données sont soit sensibles soit trop précises et évite donc l'anonymisation, ou encore, comme pour la valeur de la résidence, qui ne présente pas un grand intérêt car elles sont basées sur une estimation client (beaucoup de lignes ne sont soit pas remplir soit mise à 0).
J'ai utilisé la requête suivante :

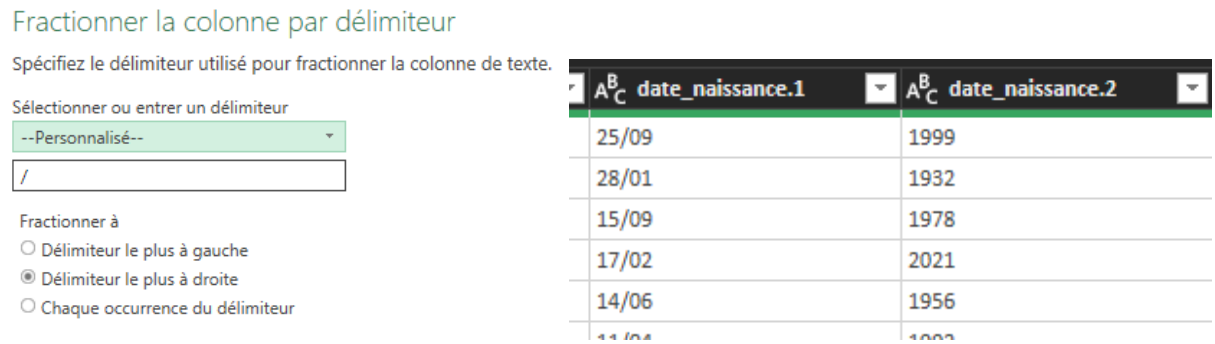


J'enregistre ensuite le résultat de cette requête dans un fichier csv.

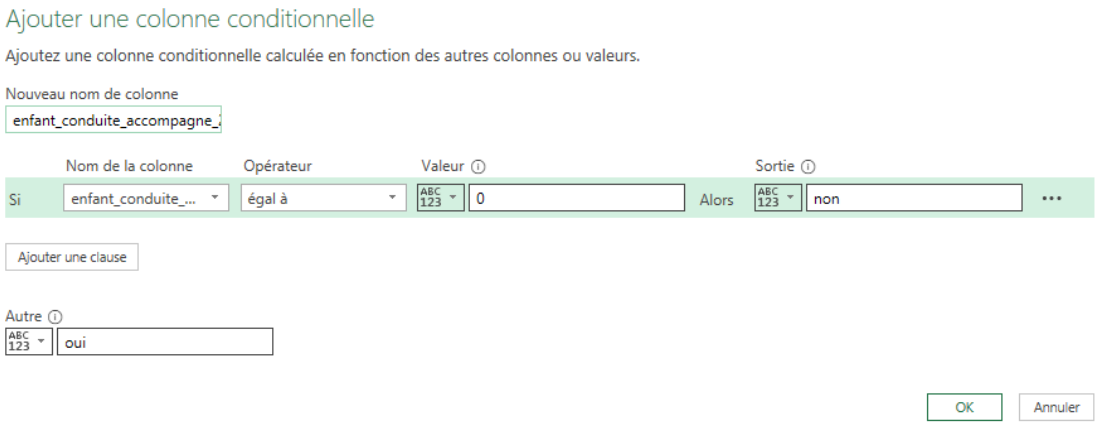
-Documentation Power Query :

Ensuite, dans Excel, dans l'onglet données je vais dans obtenir des données > à partir d'un fichier texte/CSV, pour charger le fichier de l'étape précédente, ce qui crée automatiquement une requête et connexion avec Power Query.

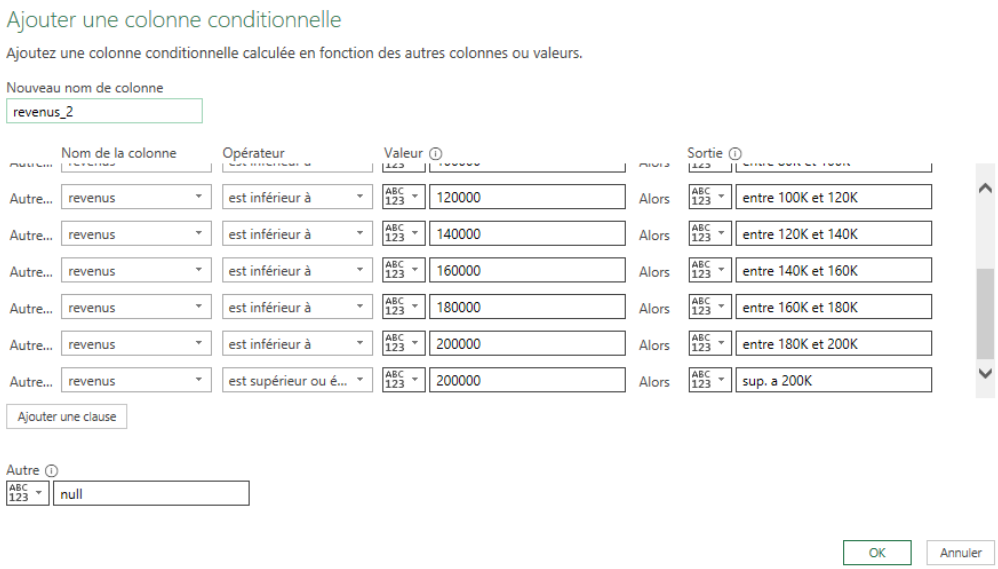
C'est à ce stade que je vais anonymiser la base de données.
Je commence par conserver uniquement l'année de naissance des personnes afin de limiter la réidentification. Pour ce faire, je fractionne la colonne par délimitateur, qui est « / », je choisi celui le plus a droite. Cela me donne deux colonnes. Je peux supprimer celle de gauche et renommer et garder celle de droite. (cf. image suivante)



Ensuite, concernant le nombre d'enfants et le nombre d'enfants en conduite accompagnée, je vais procéder de la même manière. Toujours dans le but de limiter la réidentification. Je traduis la valeur donnée par une condition : si elle est égale à 0 (ce qui correspond aux clients n'ayant pas d'enfants) cela se traduit par un « non » sinon par un « oui ». (Cf image suivante)



Concernant les revenus des clients, j'ai choisi de créer des intervalles de valeurs afin de poursuivre l'anonymisation. Au lieu d'avoir une valeur précise, j'aurais un intervalle de revenus. J'ai opté pour les intervalles suivants : inf. à 20K, entre 20K et 40K, entre 40K et 60K, entre 60K et 80K, etc... jusqu'à sup. à 200K.
Pour cela j'ai également utilisé la création d'une colonne conditionnelle (Cf image suivante). Cependant, le résultat obtenu comporte des erreurs. Cela s'explique car certaines lignes ne sont pas renseignées. J'ai donc demandé à Power Query de remplacer les valeurs erronées par des valeurs 'null'.



Je poursuis avec l'adresse du client en ne conservant que le nom de la ville et plus l'adresse complète. Pour ce faire, ce commence par fractionner la colonne avec le délimitateurs 'espace' le plus a droite. Comme certaine ville ont une préposition « la/le » que je souhaite conserver, je continuer avec le fractionnement de la première colonne selon les transitions de chiffre à non-chiffre (résultat : Cf image). Ensuite, je fusionne la colonnes 'adresse.1.4' avec 'adresse.2'. Je supprime les espaces en premier caractère (qui aurait pu se mettre du a la fusion des colonnes) et nettoie la colonne.
Je supprime les autres colonnes non nécessaires et renomme celle que je garde.

adresse.1.1	adresse.1.2	adresse.1.3	adresse.1.4	adresse.2	
87	Rue des Chauvelles 58000	*null*	*null*	Nevers	oui
10	Rue des Grands Champs 73460	*null*	*null*	Frontenex	no
30	Quai du Canal 30800	*null*	*null*	Saint-Gilles	no
29	Rue de Rigaudou 47510	*null*	*null*	Foulayronnes	no
1041	D 216	La Seyne A Six Fours 83500	La	Seyne-sur-Mer	oui
	null	*null*	*null*	*null*	no
6	Rue du Rempart 23800	*null*	*null*	Dun-le-Palestel	no

Je termine avec la création d'un index avec ajouter une colonne > colonne d'index > à partir de 1. Je lui rajoute ensuite un préfixe de six 0 avec l'onglet format > ajouter un préfixe, puis j'extrait les derniers caractères dans l'onglet extraire > derniers caractères, où je conserve uniquement 6 chiffres. Cela me donne un index à 6 chiffres.

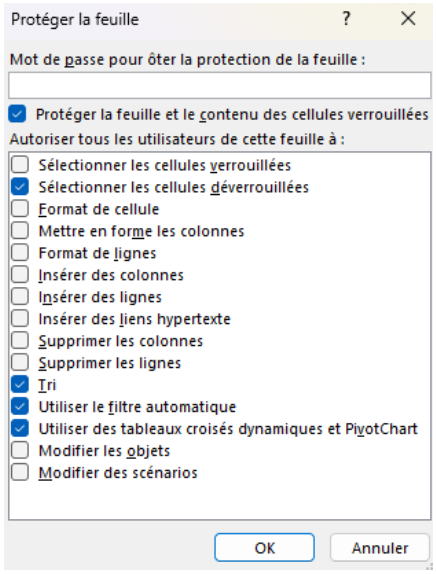
Pour la colonne 'date_demande' je modifie uniquement le type de valeur afin de conserver qu'une date en jour/mois/année et ainsi exclure l'heure de la demande.

DEV'IMMEDIAT

Suite à ces traitements effectués avec Power Query, la réidentification est désormais plus complexe, ce qui correspond à notre objectif.

-Documentation données brutes :

Le seul traitement que j'effectue dans le CSV est le masquage de la colonne 'tarif_devis'. Pour cela, je procède en plusieurs étapes. Premièrement, après avoir sélectionné l'ensemble du tableau avec Ctrl+A, dans le menu contextuel, je choisis 'format de cellule…', puis dans l'onglet protection je décoche les options 'verrouillée' et 'masquée'.
Ensuite, je sélection entièrement la colonne tarif et dans le même onglet je coche les deux options. Je masque ensuite la colonne. Dans l'onglet révision, je choisi 'protéger la feuille'. Je sélectionne certaines option (Cf image) pour permettre à l'équipe performance de travailler sur ces données. Enfin, je défini un mot de passe. L'équipe performance n'aura pas accès à ce mot de passe (seul l'équipe commercial le possédera) et donc ne pourra pas accéder aux tarifs même s'ils sont bien présents dans le document.



-Retravailler le jeu de données :

En plus des préconisations faites pour respecter le RGPD, telles que la minimisation des données, je conseille la création de volets déroulant là où c'est possible (ex : enfants ? oui ou non, revenus ? liste déroulante avec les options intervalles crées, …). Cela permettra d'avoir une base de données plus saine et ainsi de passer moins de temps sur le nettoyage des données.

DEV'IMMEDIAT

Conclusion

Il est indéniable que le respect des règles RGPD est non seulement une obligation légale, mais aussi une nécessité éthique pour toute organisation traitant des données personnelles. Il est essentiel d'adopter une approche proactive telle que la création d'une équipe dédier comme vous l'avais décidé.
La confidentialité des données n'est pas simplement une question de conformité légale, mais plutôt une composante de la confiance entre les organisations et leurs utilisateurs. Cela signifie être transparent sur la manière dont les données sont collectées, utilisées et protégées, tout en garantissant le respect des droits des individus en matière de confidentialité.

Il est nécessaire de reconnaître les contraintes et défis rencontrés dans la mise en œuvre des mesures de protection des données. Ces contraintes peuvent être techniques, financières ou organisationnelles, mais il est essentiel de les aborder de manière proactive pour trouver des solutions qui concilient efficacement les impératifs de confidentialité avec les objectifs commerciaux et opérationnels.
Il faut garder également en mémoire que la protection des données est un processus continu et évolutif. Les menaces à la confidentialité évoluent constamment, tout comme les attentes des utilisateurs en matière de protection de leur vie privée. Par conséquent, les organisations doivent rester vigilantes, flexibles et engagées à améliorer en permanence leurs pratiques de protection des données. Ce que faciliteras l'équipe créer.

Les enjeux de la confidentialité sont cruciaux tant sur le plan éthique que commercial. Les entreprises doivent non seulement respecter les règles RGPD, mais elles peuvent également renforcer la confiance des utilisateurs pour maintenir des relations durables avec la clientèle, réduire les risques juridiques et opérationnels (les violations de la confidentialité peuvent non seulement entraîner des sanctions financières importantes, mais aussi nuire à la réputation et à la crédibilité d'une entreprise), et positionner leur entreprise comme un leader en matière de protection des données.