

Modèle de rapport de traitement des données

Titre du projet	Collectez des données en respectant les normes RGPD
-----------------	---

Version	Auteur	Description	Date
V1	Wafa Zargouni	Rapport sur les traitements des données de Dev'Immédiat.	08/11/2023

Introduction

L'union européenne a mis en place un règlement de protection des données à caractère personnel l'RGPD (Règlement général sur la protection des Données) et a créé une autorité administrative indépendante la CNIL (Commission Nationale de l'informatique et des libertés) pour assurer le respect de ce règlement. Le CNIL a un rôle d'alerte, de conseil et d'information vers tous les publics mais dispose également d'un pouvoir de contrôle et de sanction.

Cependant, Dev'immediat, le courtier d'assurance auto, vient d'être sanctionné par l'interdiction d'accès aux données personnelles de ces clients temporairement pendant 6 mois à cause d'une alerte faite par un de ses clients au CNIL qui avait soupçonné l'entreprise de traiter illégalement ses données personnelles.

Cette interdiction d'accès aux données personnelles bloque le bon fonctionnement de l'entreprise puisque le traitement des données personnelles de ses clients est au cœur de son activité professionnelle. Pour cela, l'entreprise nécessite en urgence une solution temporaire à utiliser pendant cette période de sanction qui respecte les règles de protection des données. Cette solution consiste à fournir une extraction des données anonymisées.

Solution temporaire

La solution temporaire consiste à produire une extraction des données anonymisées des demandes faites en 2022. Cette extraction des données anonymisées permettra à la fois : de rester conforme avec la sanction puisque on ne traite plus des données personnelles.

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible.

Les principaux techniques de l'anonymisation de données sont :

- Randomisation : consiste à modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale. Cette technique permet de protéger le jeu de données du risque d'inférence

- Généralisation : consiste à modifier l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes. Cette technique permet d'éviter l'individualisation d'un jeu de données. Elle limite également les possibles corrélations du jeu de données avec d'autres.

L'anonymisation des données doit respecter des critères d'efficacité telles que :

- l'individualisation : il ne doit pas être possible d'isoler un individu dans le jeu de données ;
- la corrélation : il ne doit pas être possible de relier entre eux des ensembles de données distincts concernant un même individu ;
- l'inférence : il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

Les étapes de traitement des données à effectuer :

1. Extraction des données brutes de la base des données:

- Ouvrir la base de données sur SQLite ;
- Exécuter la requête SQL pour extraire les données de l'année 2022 (la demande d'assurance a été effectuée en 2022) pour les clients dont l'état du dossier est « complet »;
- Exporter le résultat de la requête en CSV;






Requête

Historique




```
1 select date_demande, date_naissance, nombre_enfants, revenus,
2 usage_vehicule, type_vehicule, points_perdus, type_conduite, formule
3 from base_client
4 where etat_dossier = 'complet' and |
5 date_demande >= '2022-01-01 00:00:00'
6 and date_demande < '2023-01-01 00:00:00';
7
```

Table

Formulaire



1



Nombre de lignes chargées : 1158

	date demande	date naiss	nombre e	revenus	usage vehic	type vehic	points pe	type conduite	formule
1	2022-02-25 16:13:38	1999-09-25	1	125301	Commercial	Sports Car	0	Highly Urban/ Urban	dev_integral
2	2022-05-24 21:07:34	1932-01-28	1	50815	Private	Minivan	2	Highly Urban/ Urban	dev_express
3	2022-01-24 08:17:21	1978-09-15	2	21204	Private	Pickup	0	z_Highly Rural/ Rural	dev_express
4	2022-04-20 03:46:00	2021-02-17	0	0	Private	z_SUV	3	Highly Urban/ Urban	dev_integral
5	2022-03-23 01:32:40	1956-06-14	0	NULL	Commercial	Panel Truck	2	Highly Urban/ Urban	dev_vip
6	2022-03-23 22:20:50	1992-04-11	1	79606	Private	z_SUV	2	z_Highly Rural/ Rural	dev_premium
7	2022-03-13 01:10:59	1941-03-18	2	130540	Commercial	Minivan	0	z_Highly Rural/ Rural	dev_express
8	2022-03-09 03:43:53	1956-03-06	0	51278	Private	z_SUV	1	Highly Urban/ Urban	dev_premium
9	2022-01-29 10:17:49	1956-02-18	0	115044	Private	Minivan	0	Highly Urban/ Urban	dev_premium

Figure 1 : Résultat de la requête SQL

	A	B	C	D	E	F	G	H	I	
	date_demande	date_naissance	nombre_enfants	revenus	usage_vehicule	type_vehicule	points_perdus	type_conduite	formule	
1	25/02/2022 16:13	25/09/1999	1	125301	Commercial	Sports Car	0	Highly Urban/ Urb	dev_integral	
2	24/05/2022 21:07	28/01/1932	1	50815	Private	Minivan	2	Highly Urban/ Urb	dev_express	
3	24/01/2022 08:17	15/09/1978	2	21204	Private	Pickup	0	z_Highly Rural/ Ru	dev_express	
4	20/04/2022 03:46	17/02/2021	0	0	Private	z_SUV	3	Highly Urban/ Urb	dev_integral	
5	23/03/2022 01:32	14/06/1956	0		Commercial	Panel Truck	2	Highly Urban/ Urb	dev_vip	
6	23/03/2022 22:20	11/04/1992	1	79606	Private	z_SUV	2	z_Highly Rural/ Ru	dev_premium	
7	13/03/2022 01:10	18/03/1941	2	130540	Commercial	Minivan	0	z_Highly Rural/ Ru	dev_express	
8	09/03/2022 03:43	06/03/1956	0	51278	Private	z_SUV	1	Highly Urban/ Urb	dev_premium	
9	29/01/2022 10:17	18/02/1956	0	115044	Private	Minivan	0	Highly Urban/ Urb	dev_premium	
10	26/03/2022 07:37	07/03/1974	0		Commercial	Panel Truck	1	Highly Urban/ Urb	dev_integral	
11	26/05/2022 05:51	13/02/1967	2	24060	Private	Minivan	0	Highly Urban/ Urb	dev_express	
12	27/01/2022 09:33	14/05/2017	4	20822	Private	z_SUV	0	Highly Urban/ Urb	dev_premium	
13	08/05/2022 13:32	20/12/1950	0	74426	Private	z_SUV	0	Highly Urban/ Urb	dev_express	
14	02/05/2022 08:44	26/07/2017	0	0	Commercial	Van	5	z_Highly Rural/ Ru	dev_integral	
15	15/04/2022 13:09	07/10/1945	1	38042	Private	Minivan	0	Highly Urban/ Urb	dev_integral	
16	11/05/2022 02:24	12/02/1938	0	45787	Private	Minivan	0	Highly Urban/ Urb	dev_express	

Figure 2: données brutes exportées en CSV

2. Traitement des données sur Power Query:

- Importer les données sur Power BI ;
- Ouvrir la fenêtre de transformation des données ;
- Ajouter une colonne index : Cette colonne est utilisée comme un identifiant. Elle contient des nombres arbitraires;
- Modifier le type de la colonne « date_naissance » en Date ;
- Renommer la colonne « date_naissance » à « age » et le calculer en année ;
- Créer une nouvelle colonne conditionnelle « tranches_age » à partir de la colonne « date_naissance » qui contient les trois valeurs : « agé », « adulte » ou « jeune » selon les conditions suivantes :

 si [age] >= 65 alors "agé"
 sinon si [age] > 35 alors "adulte"
 sinon si [age] <= 35 alors "jeune"
 sinon null
- Modifier le type de la colonne « revenus » et remplacer les valeurs par 0 ;
- Créer une nouvelles colonne conditionnelle « classes_revenu » à partir de la colonne « revenus » qui contient les valeurs : ... selon les conditions suivantes :

si [revenus] > 100000 then "+100K"
sinon si [revenus] > 75000 then "[+75K,100K]"
sinon si [revenus] > 50000 then "[+50K,75K]"
sinon si [revenus] > 25000 alors "[+25K,50K]"
sinon si [revenus] <= 25000 alors "[0,25K]"
sinon null

- Modifier le type de la colonne « nombre_enfants » à entier ;
- Créer une nouvelle colonne conditionnelle « a_enfants » à partir de la colonne « nombre_enfants » qui contient les deux valeurs « oui » si le client a au moins un enfant ou « Non » s'il n'y en a pas.
- Masquer les colonnes « age », « revenus » et « nombre_enfants »
- Créer une visualisation des données sous forme tableau et exporter les données traitées sous format CSV.

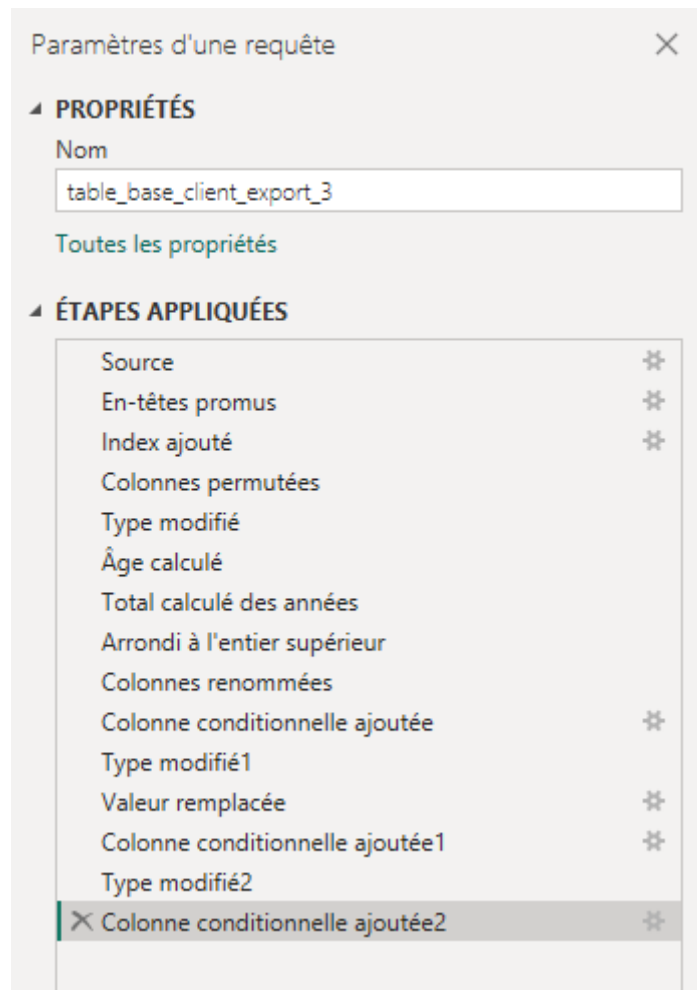


Figure 3: les étapes de traitement appliquées sur Power Query

A	B	C	D	E	F	G	H	I	J
Index	date_demande	formule	points_perdus	tranches_age	classes_revenu	a_enfants	type_conduite	type_vehicule	usage_vehicule
0	25/02/2022 16:13	dev_integral	0	jeune	+100K	Oui	Highly Urban/ U	Sports Car	Commercial
1	24/05/2022 21:07	dev_express	2	agé	[+50K;75K]	Oui	Highly Urban/ U	Minivan	Private
2	24/01/2022 08:17	dev_express	0	adulte	[0;25K]	Oui	z_Highly Rural/	Pickup	Private
3	20/04/2022 03:46	dev_integral	3	jeune	[0;25K]	Non	Highly Urban/ U	z_SUV	Private
4	23/03/2022 01:32	dev_vip	2	agé	[0;25K]	Non	Highly Urban/ U	Panel Truck	Commercial
5	23/03/2022 22:20	dev_premium	2	jeune	[+75K;100K]	Oui	z_Highly Rural/	z_SUV	Private
6	13/03/2022 01:10	dev_express	0	agé	+100K	Oui	z_Highly Rural/	Minivan	Commercial
7	09/03/2022 03:43	dev_premium	1	agé	[+50K;75K]	Non	Highly Urban/ U	z_SUV	Private
8	29/01/2022 10:17	dev_premium	0	agé	+100K	Non	Highly Urban/ U	Minivan	Private
9	26/03/2022 07:37	dev_integral	1	adulte	[0;25K]	Non	Highly Urban/ U	Panel Truck	Commercial
10	26/05/2022 05:51	dev_express	0	adulte	[0;25K]	Oui	Highly Urban/ U	Minivan	Private
11	27/01/2022 09:33	dev_premium	0	jeune	[0;25K]	Oui	Highly Urban/ U	z_SUV	Private
12	08/05/2022 13:32	dev_express	0	agé	[+50K;75K]	Non	Highly Urban/ U	z_SUV	Private
13	02/05/2022 08:44	dev_integral	5	jeune	[0;25K]	Non	z_Highly Rural/	Van	Commercial
14	15/04/2022 13:09	dev_integral	0	agé	[+25K;50K]	Oui	Highly Urban/ U	Minivan	Private
15	11/05/2022 02:24	dev_express	0	agé	[+25K;50K]	Non	Highly Urban/ U	Minivan	Private
16	12/04/2022 01:20	dev_integral	0	adulte	+100K	Non	Highly Urban/ U	z_SUV	Private
17	04/03/2022 09:15	dev_express	1	agé	[+50K;75K]	Non	Highly Urban/ U	Van	Commercial
18	05/02/2022 17:39	dev_express	0	agé	[+75K;100K]	Non	Highly Urban/ U	Minivan	Private
19	25/04/2022 14:43	dev_integral	0	agé	[+25K;50K]	Non	Highly Urban/ U	z_SUV	Private
20	20/03/2022 03:44	dev_express	0	adulte	[+50K;75K]	Oui	z_Highly Rural/	Minivan	Private
21	27/03/2022 09:22	dev_premium	1	agé	[+25K;50K]	Oui	Highly Urban/ U	Minivan	Private
22	16/05/2022 02:39	dev_premium	0	jeune	[0;25K]	Non	Highly Urban/ U	Van	Commercial
23	06/02/2022 07:52	dev_premium	5	agé	[+25K;50K]	Oui	Highly Urban/ U	Sports Car	Commercial

Figure 4: Résultat final de l'extraction des données anonymisées

3. Conclusion

Par le biais de ces techniques d'anonymisation on a pu construire une extraction des données ce qui permet de débloquent les équipes de l'entreprise mais on a perdu plusieurs données sensibles et utiles qu'on ne peut pas traiter pendant cette période temporaire à cause de la sanction. Pour cela, il faut mettre en place un mécanisme durable de protection des données à caractère personnel.