



TARIFICATION AUTO AU BRÉSIL

10/01/2025

Master 2 Actuariat ISUP

Présenté par :

Miguel BAENA

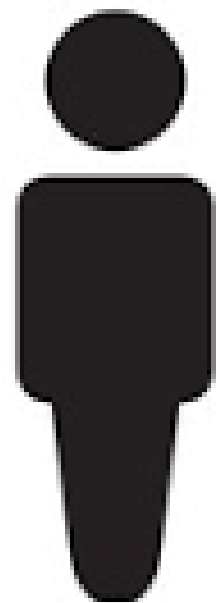
Paul DAKPOGAN

Jérémy XU



Contexte

gov.br

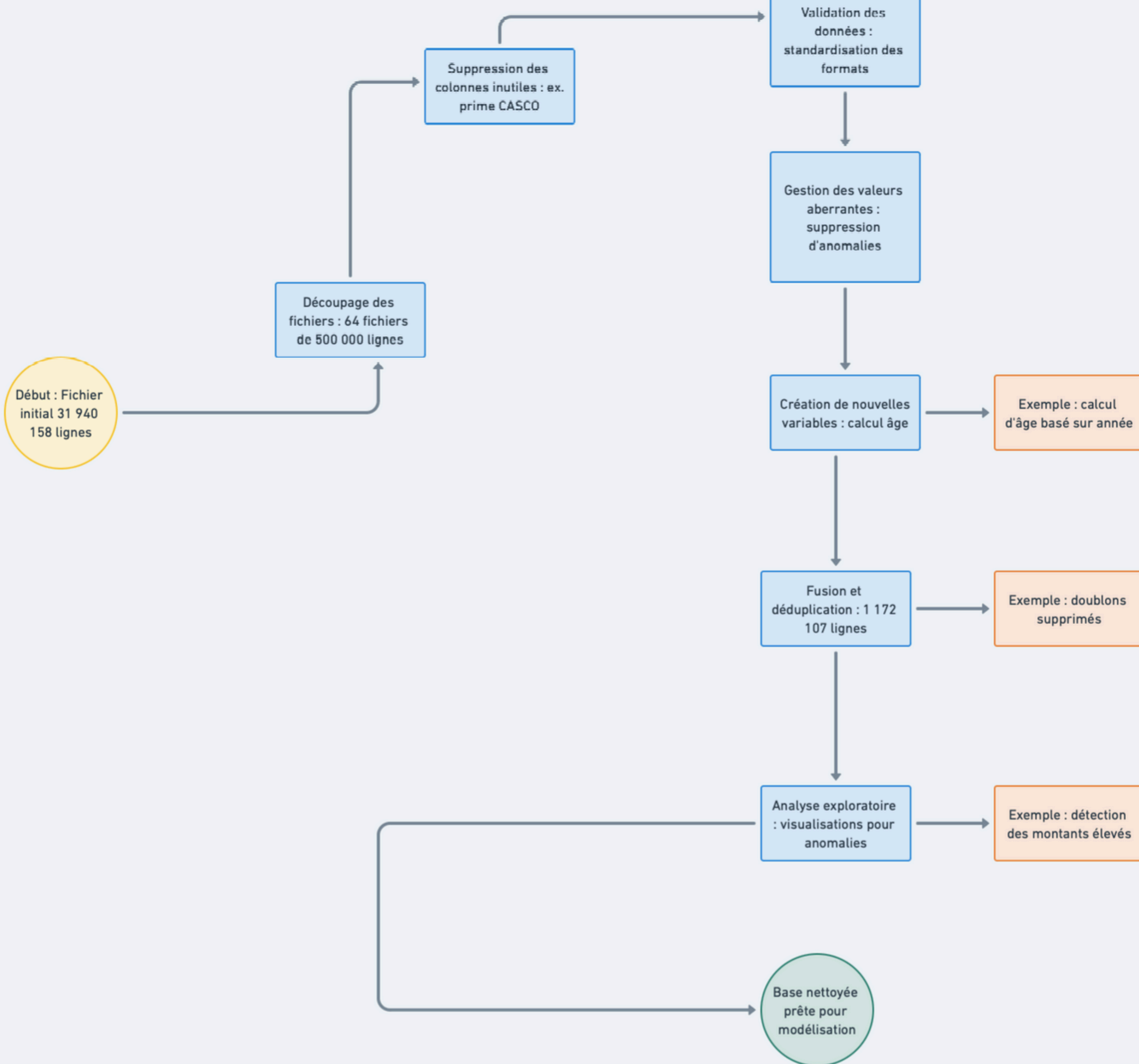


Présentation des données

Catégorie	Variable	Description
Informations liées aux accidents	REGIAO	Localisation géographique
	CEP	Code postal
	EVENTO	Type de sinistre
	CAUSA	Cause spécifique
Informations liées aux personnes	SEXO	Sexe de l'assuré
	AGE	Âge de l'assuré
	DATA_NASC	Date de naissance
	UTILIZACAO	Type d'utilisation du véhicule
Informations liées aux véhicules	COD_MODELO	Code du modèle
	ANO_MODELO	Année de fabrication
	COD_TARIF	Code tarifaire
	COBERTURA	Couverture d'assurance
	IS_RCDMAT, IS_RCDC, IS_RCDMOR	Responsabilité civile (dommages matériels, corporels, moraux)
Variables cibles	INDENIZ	Montant de l'indemnité
	number_of_claims	Nombre de sinistres déclarés



Nettoyage des données

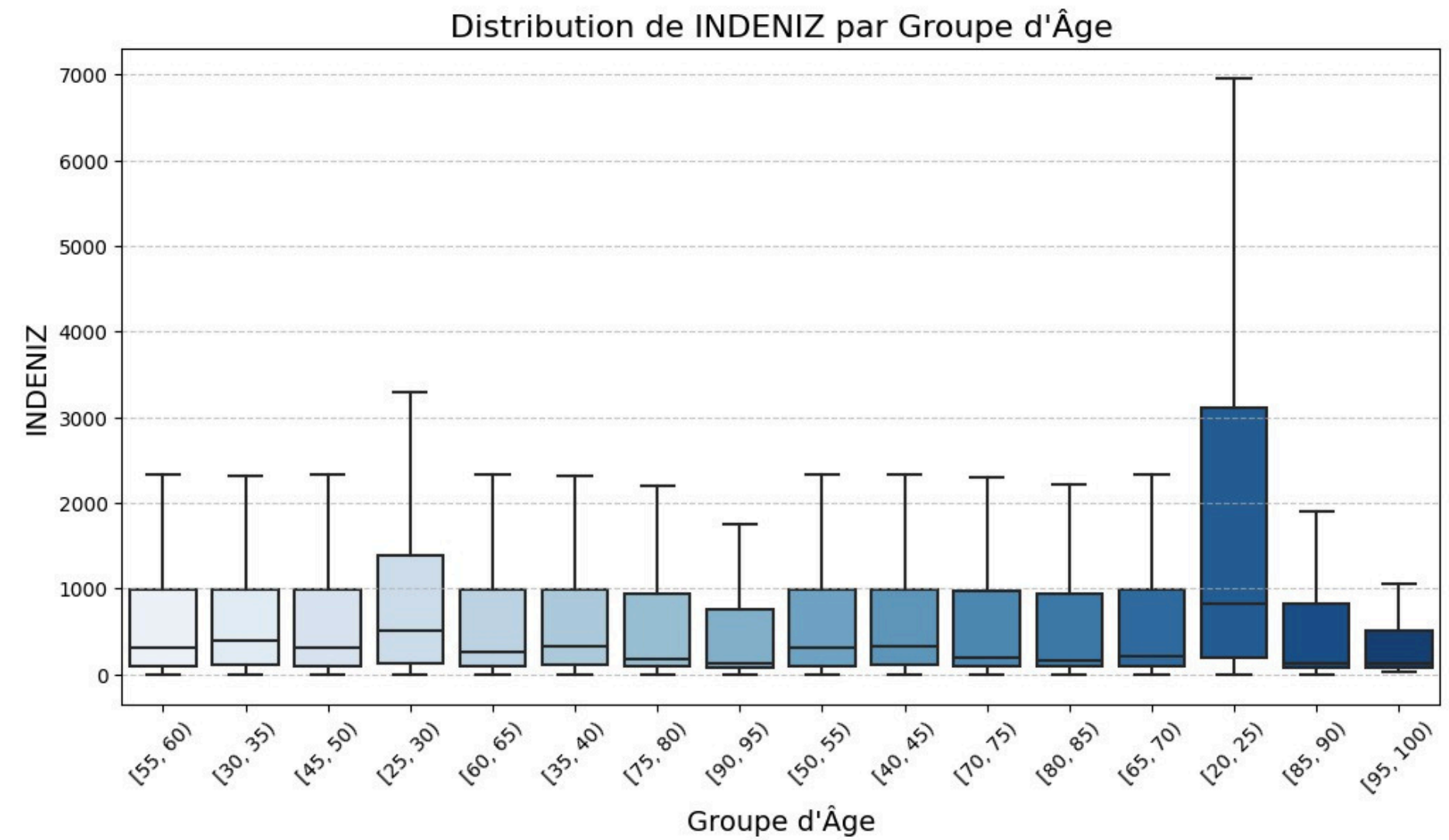
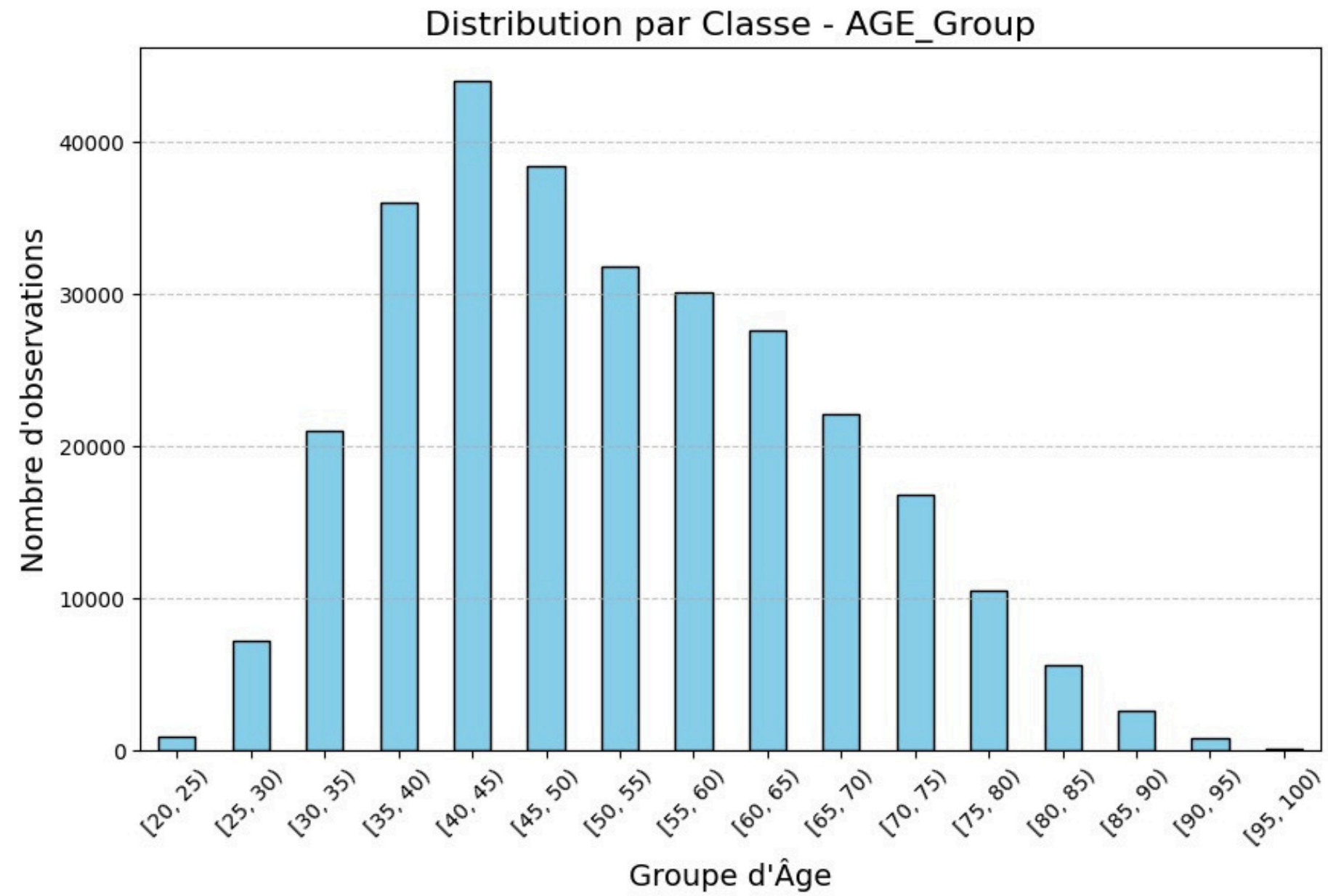


Travail effectué

Division de la Base initial en 64 fichiers qui ont été traités puis fusionnés pour obtenir une Base exploitable

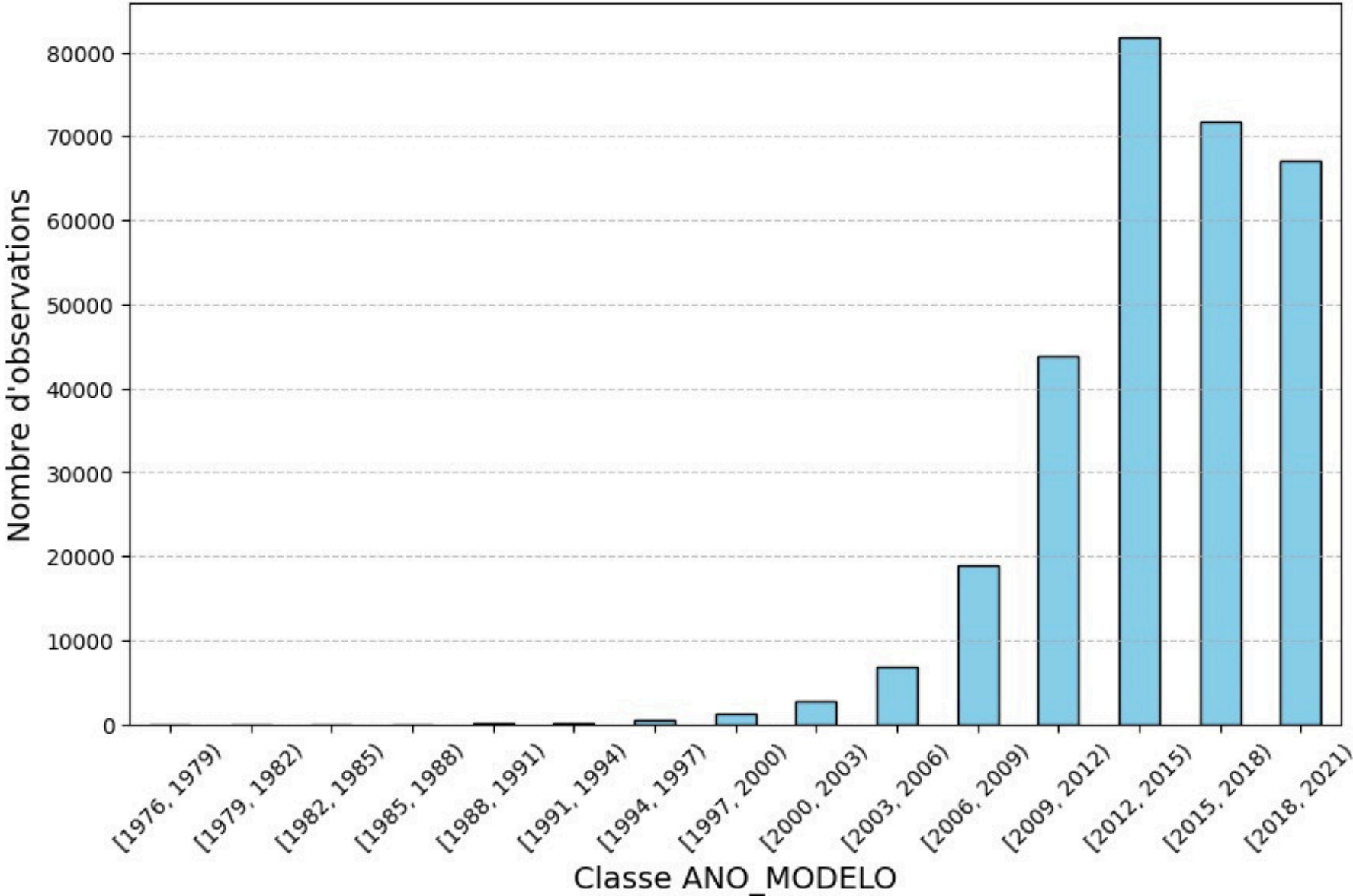


Exploration des données (AGE)

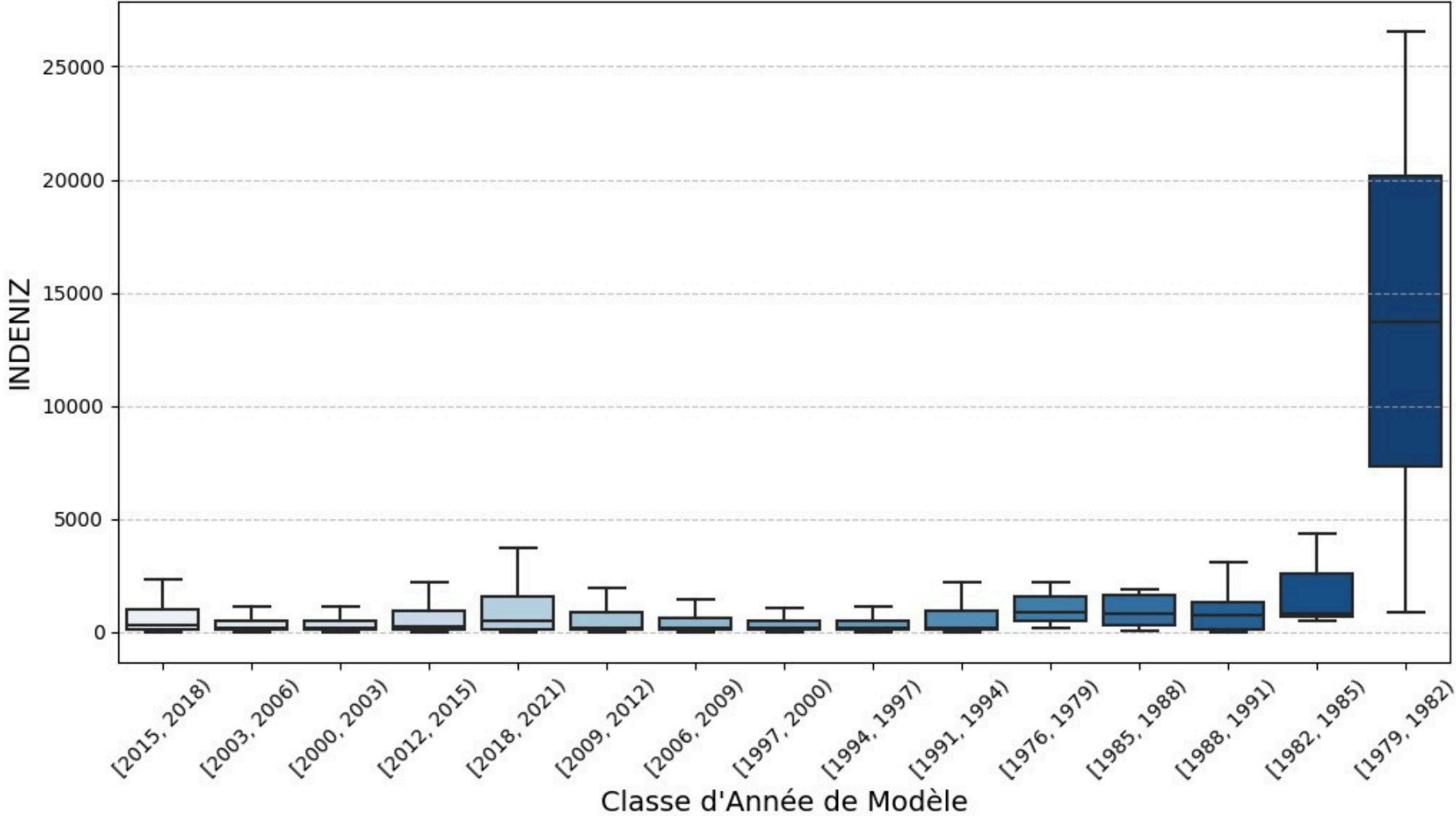


Exploration des données (Année du modèle)

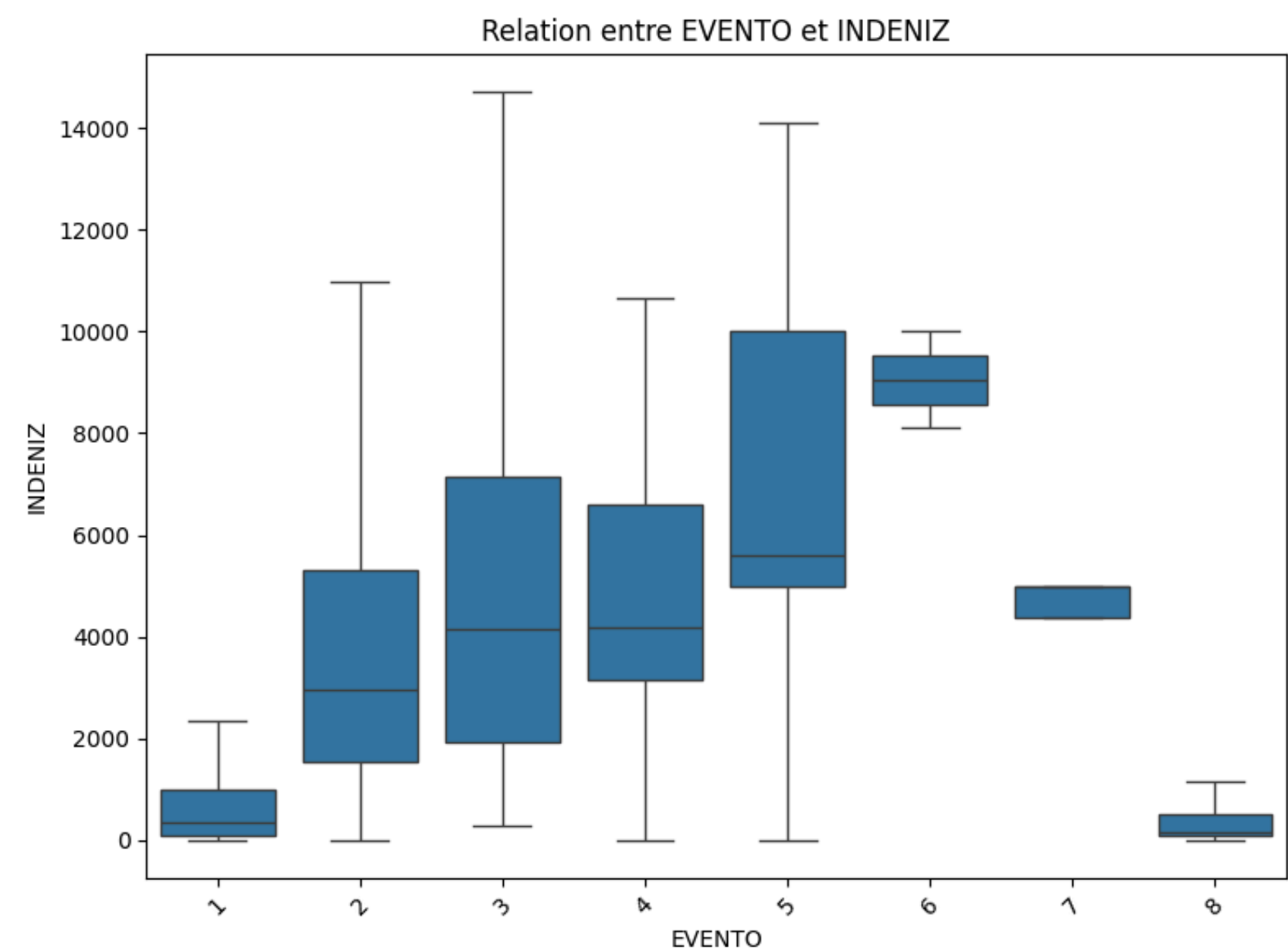
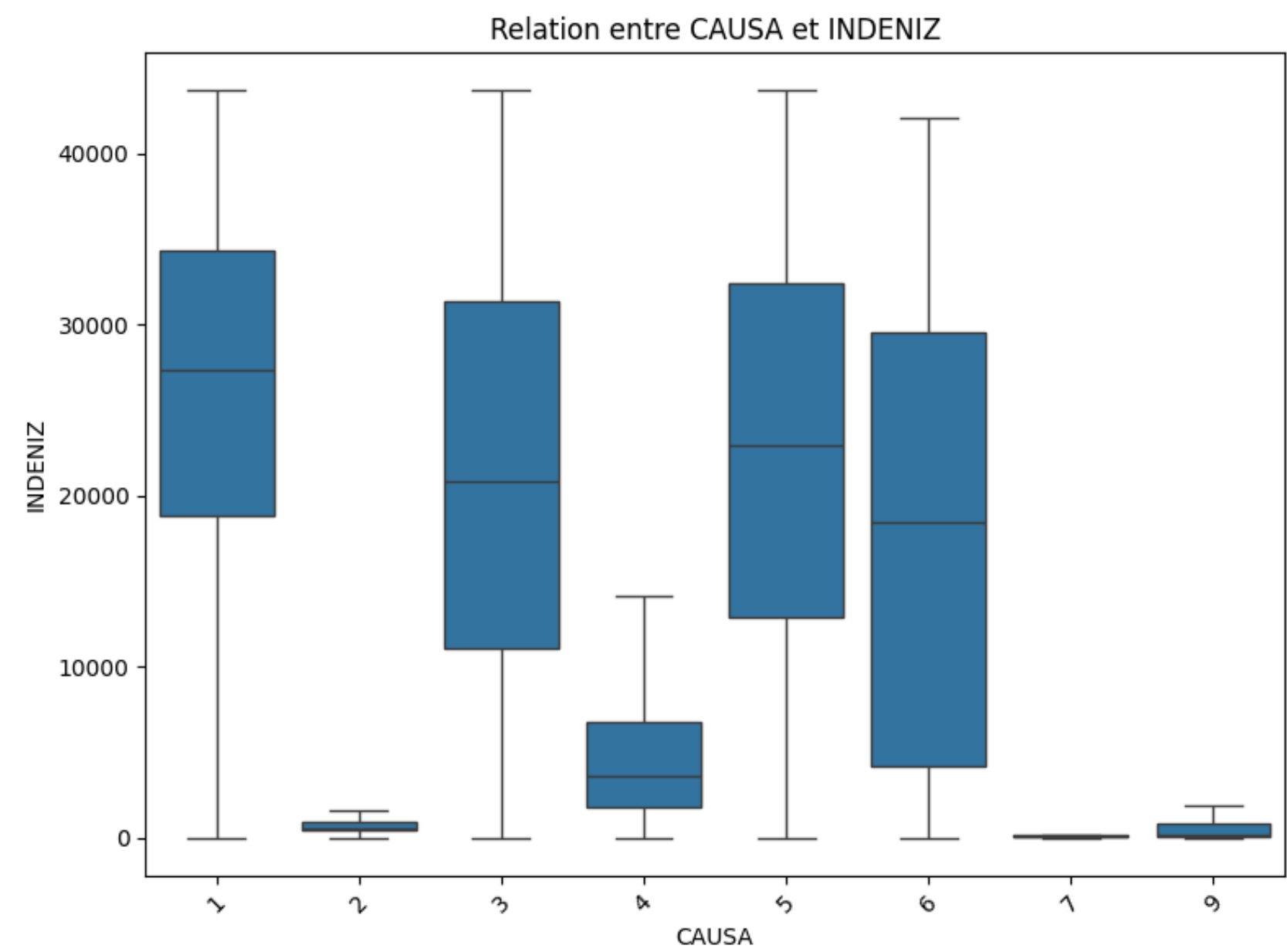
Distribution par Classe - ANO_MODELO_Class



Distribution de INDENIZ par Classe d'Année de Modèle



Exploration des données (Cause et Type de sinistre)

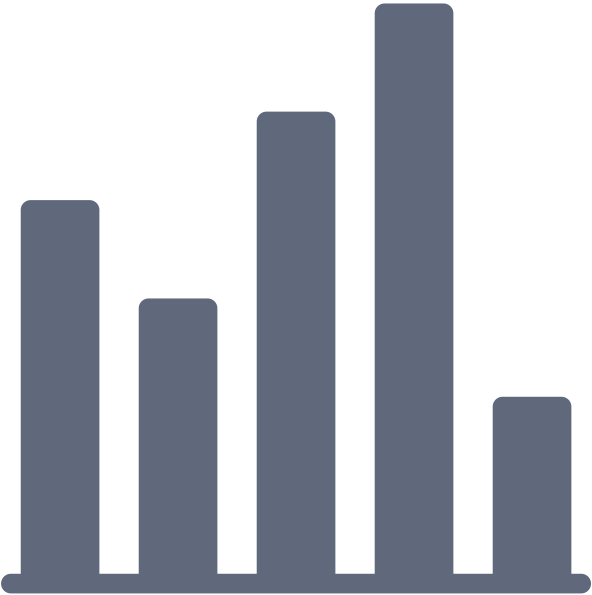


Etude de la sévérité : GLM

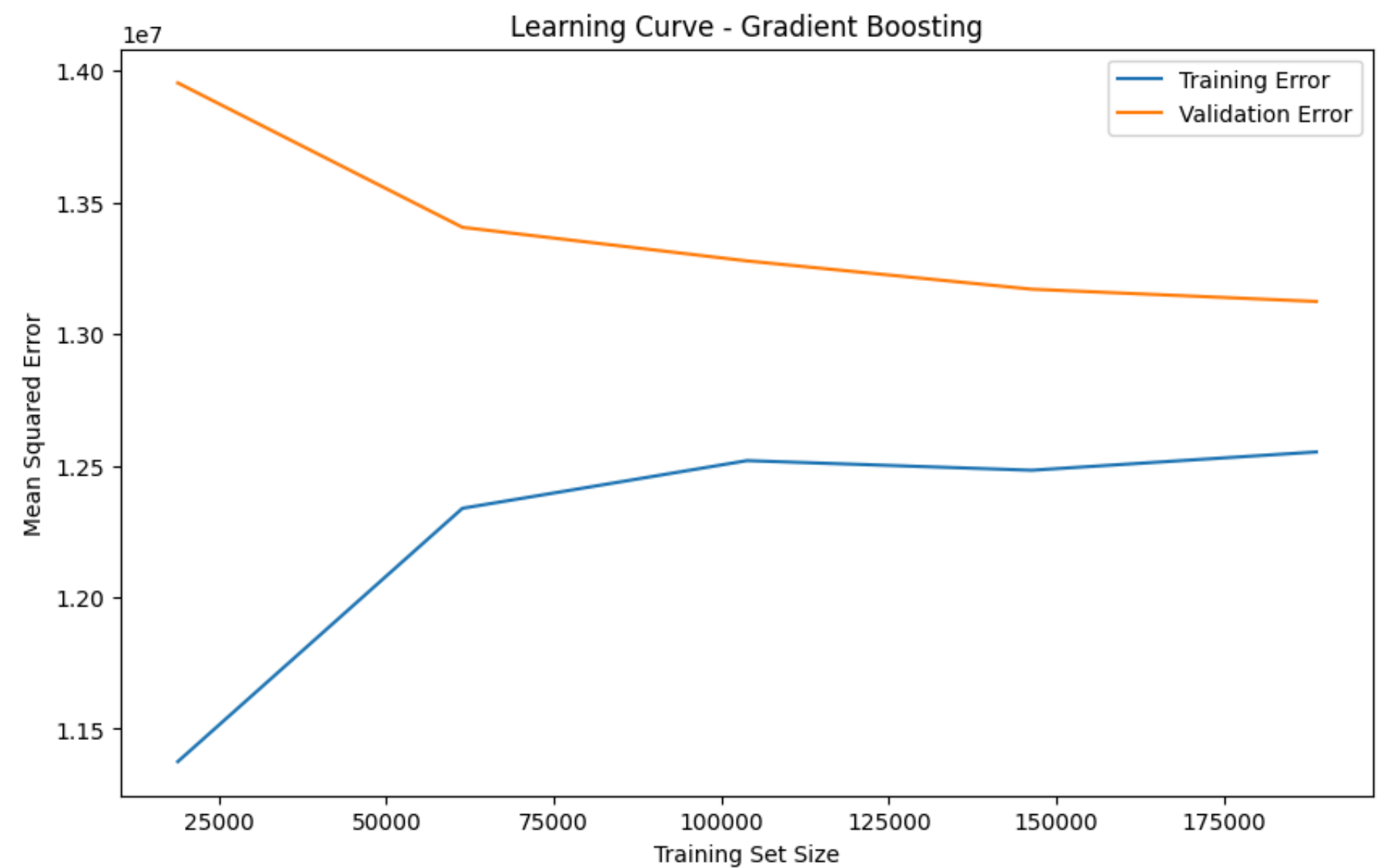
Séparation du jeu de données en train (80%) et test (20%)

Modèle : INDENIZ ~ REGIAO + SEXO + EVENTO + CAUSA

Loi	Fonction de lien	AIC
Gaussienne	Identité	1 527 894
Gausienne	Inverse	999 230
Gamma	Inverse	956 640
Gamma	Logarithme	955 860

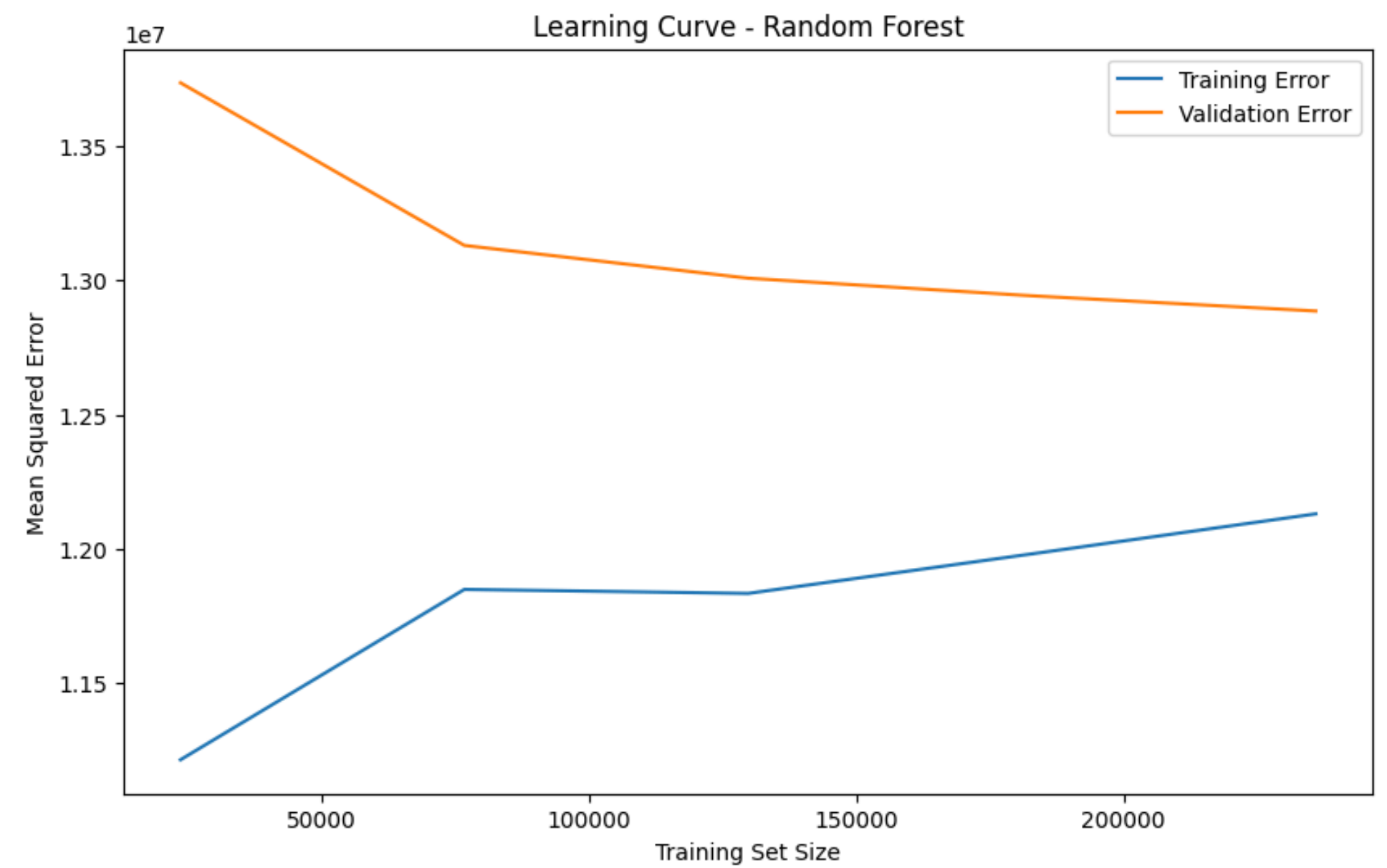


Etude de la sévérité : XGBoost



MAE	MSE	RMSE	R ²
1343.90	12746650.18	3570.25	0.49

Etude de la sévérité : Random Forest

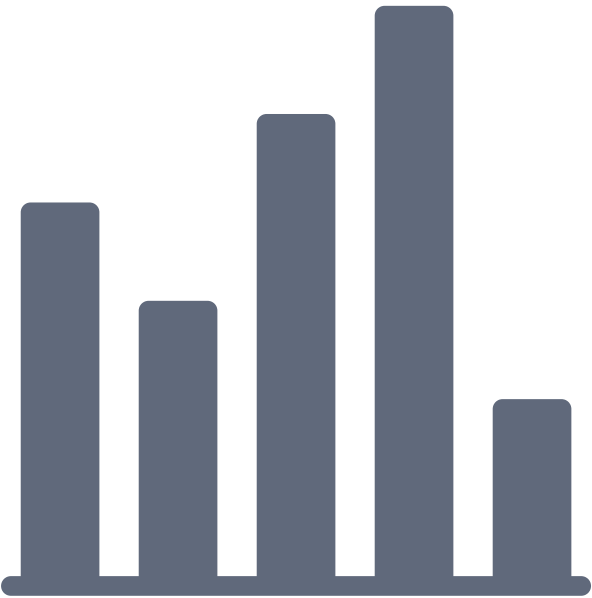


MAE	MSE	RMSE	R ²
1313.47	12603294.98	3550.11	0.50

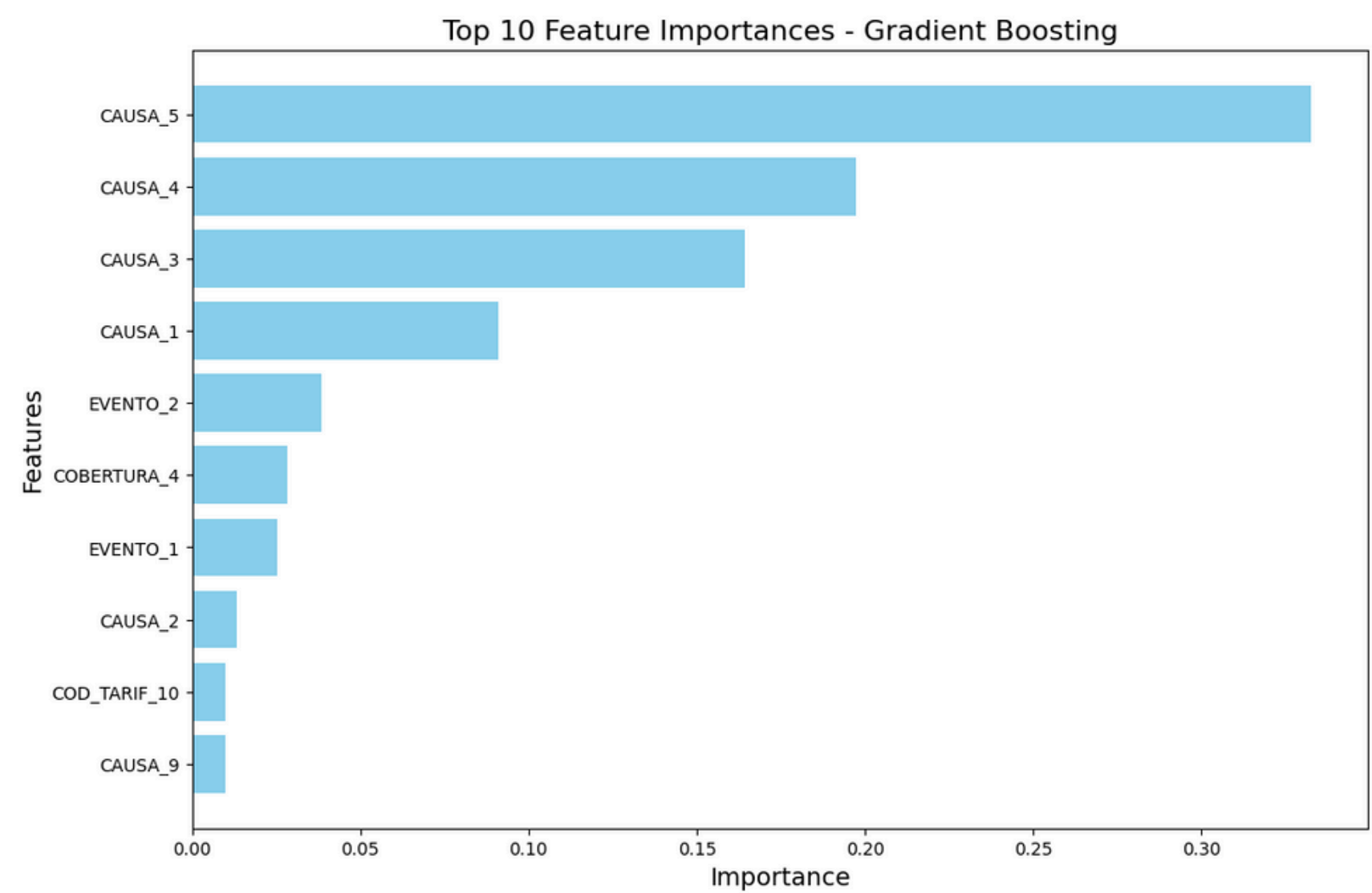
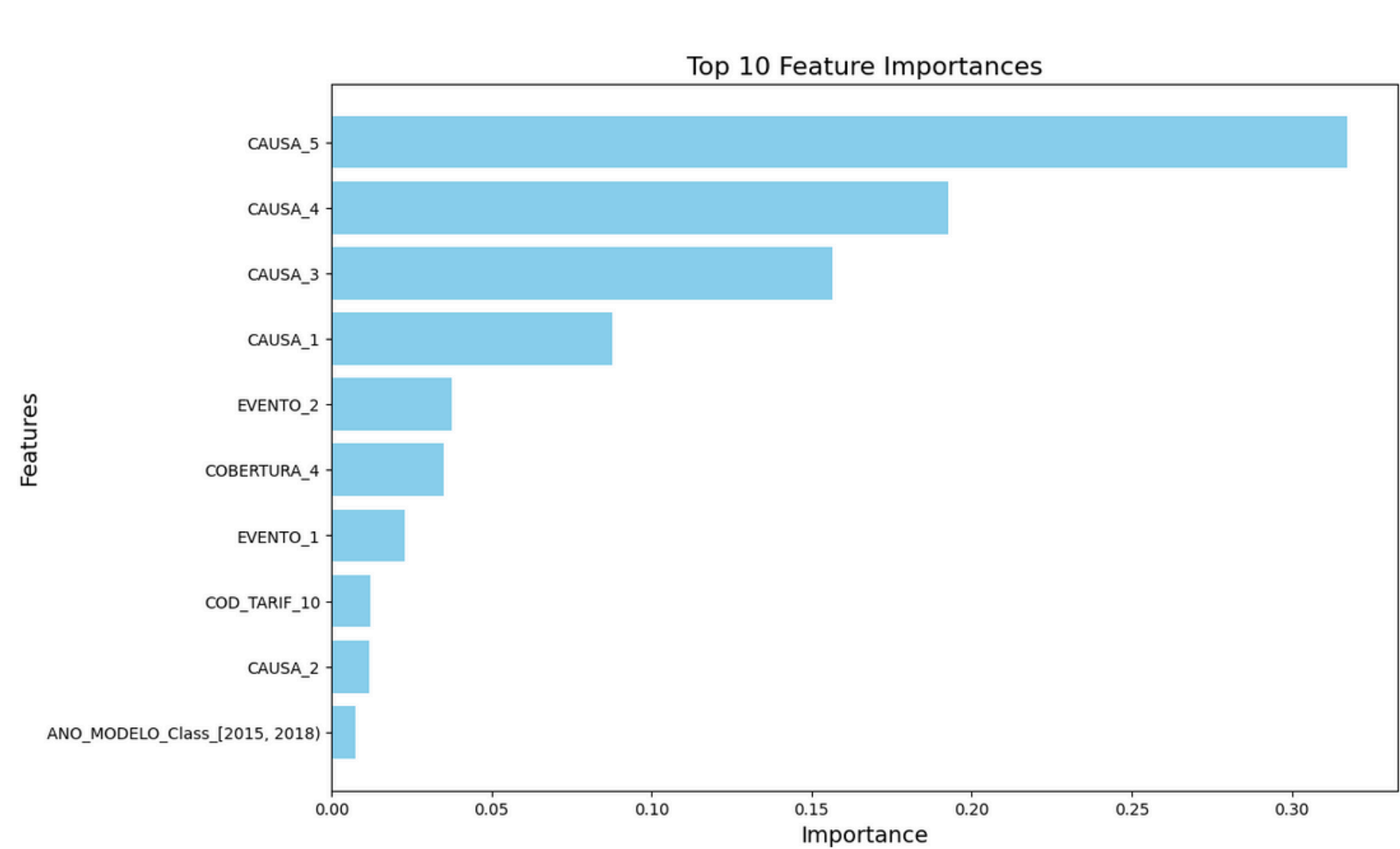
Etude de la sévérité: Comparaison des modèles

Séparation du jeu de données en train (80%) et test (20%) et dans une logique de **5-FOLD CROSS VALIDATION**

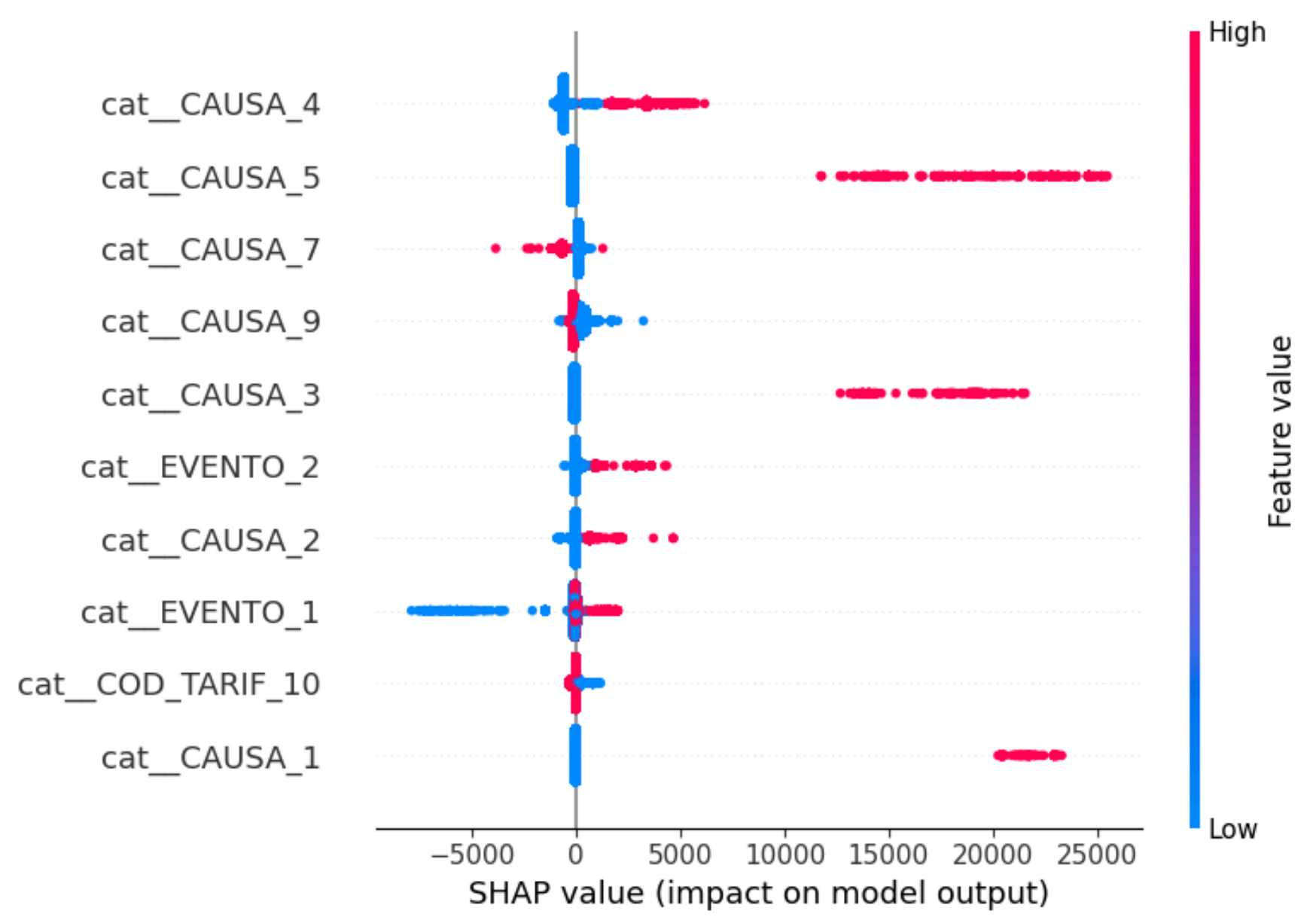
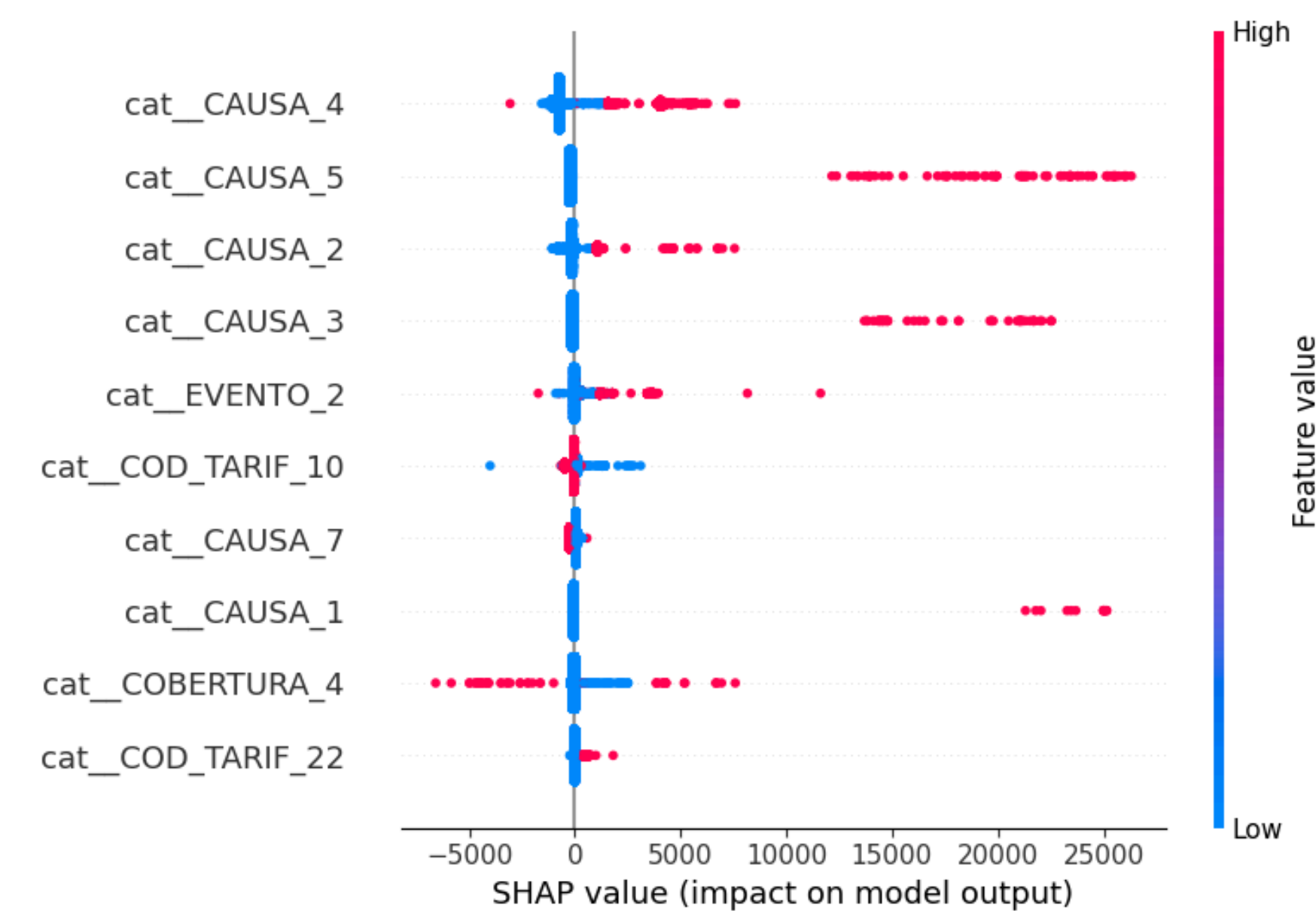
Modèle	MAE	MSE	RMSE	R ²
Random Forest	1313.47	12603294.98	3550.11	0.50
GLM Gamma	1345.32	12748321.47	3572.81	0.49
Xgboost	1343.90	12746650.18	3570.25	0.49



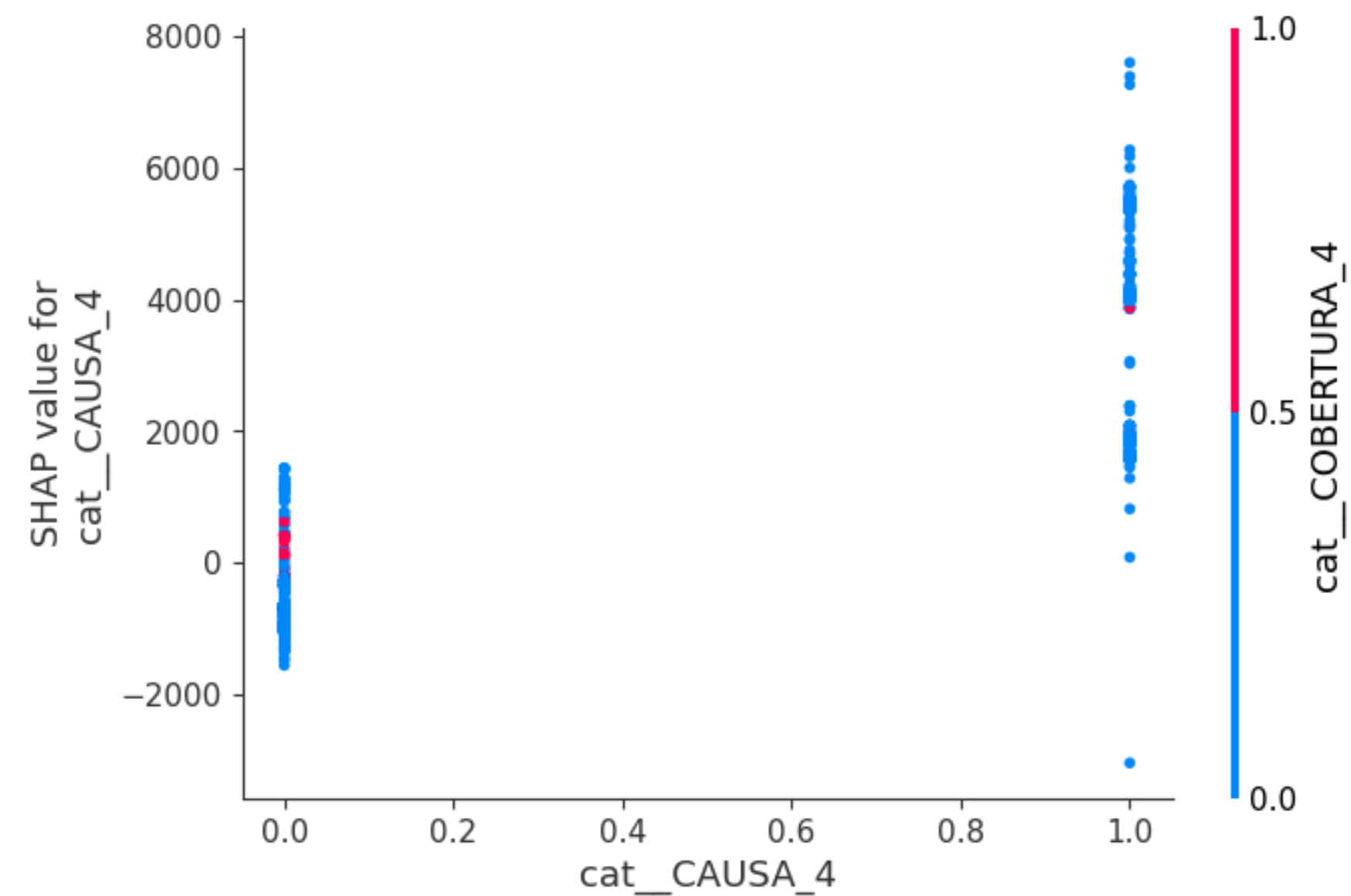
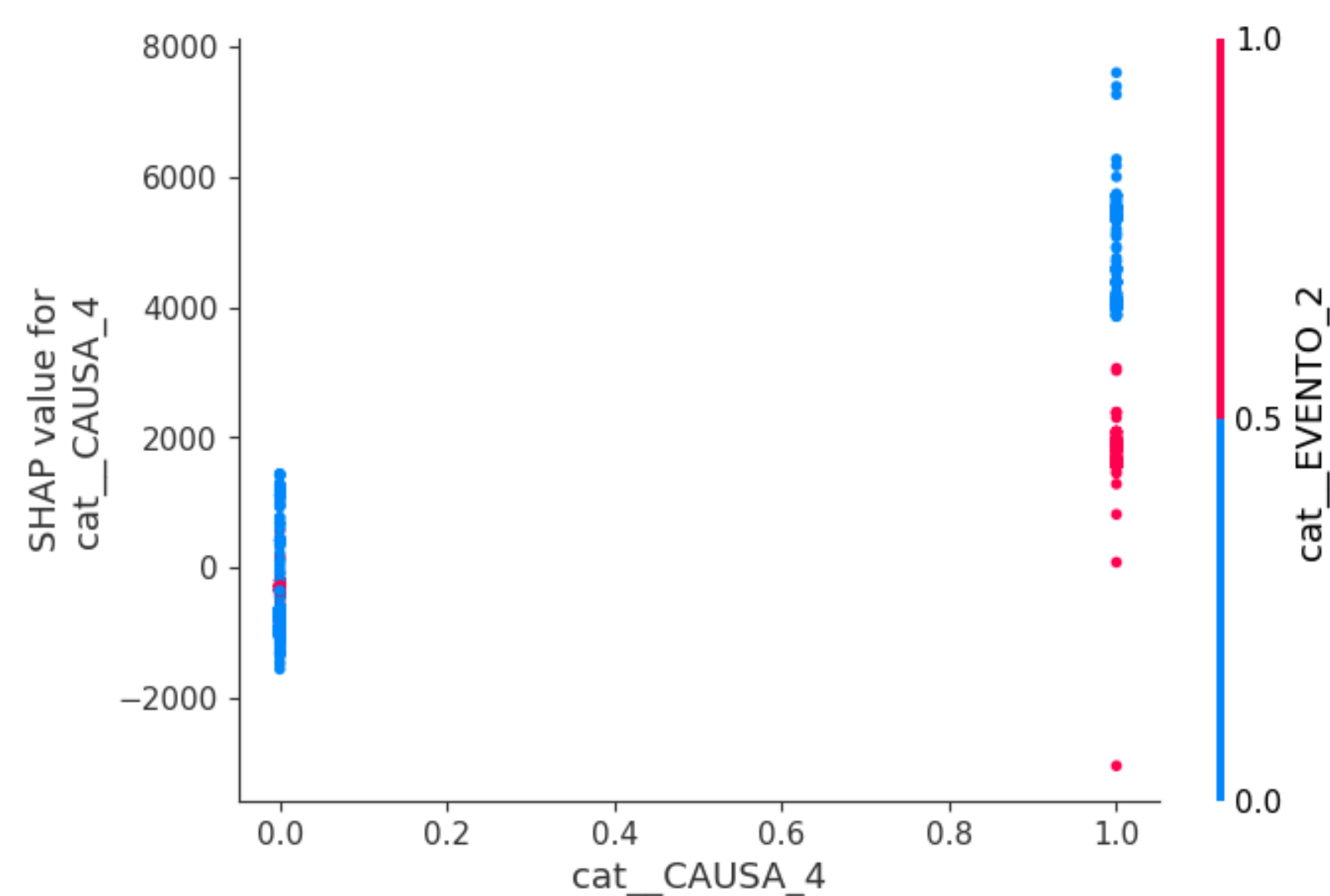
Importance des variables : RandomForest et Xgboost



Shap : RandomForest et Xgboost



Dépendance Partiel : RandomForest



Etude de la fréquence : GLM

Modèle : number of claims ~ COD TARIF_x + REGIAO_x + AGE_x + IS RCDMAT

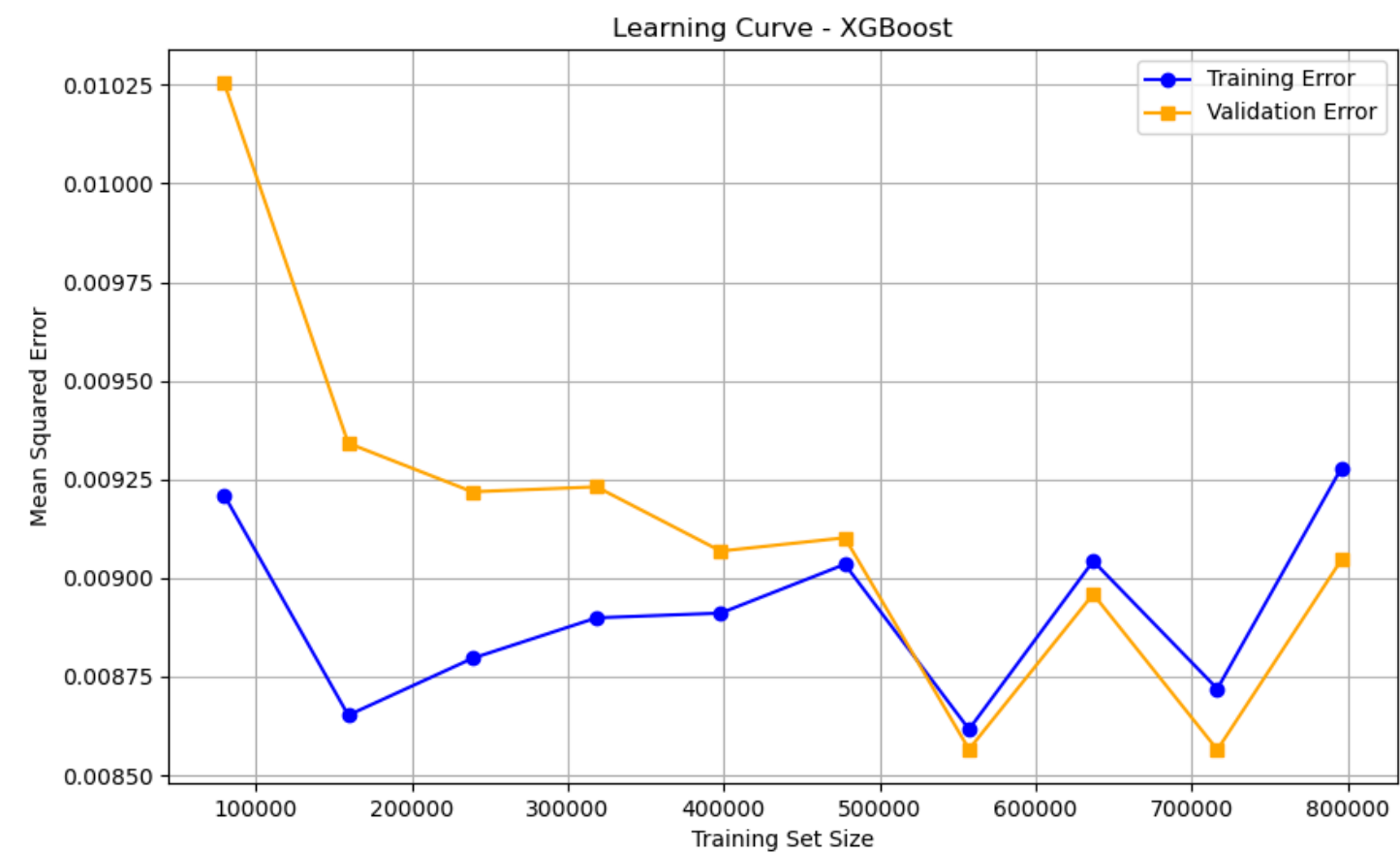
Loi	Fonction de lien	AIC
Binomiale négative	Logarithme	659 609
Poisson	Logit	174 779



Var(N)	0.44
E(N)	1.8557
Var(N)/E(N)	1.1346 > 1

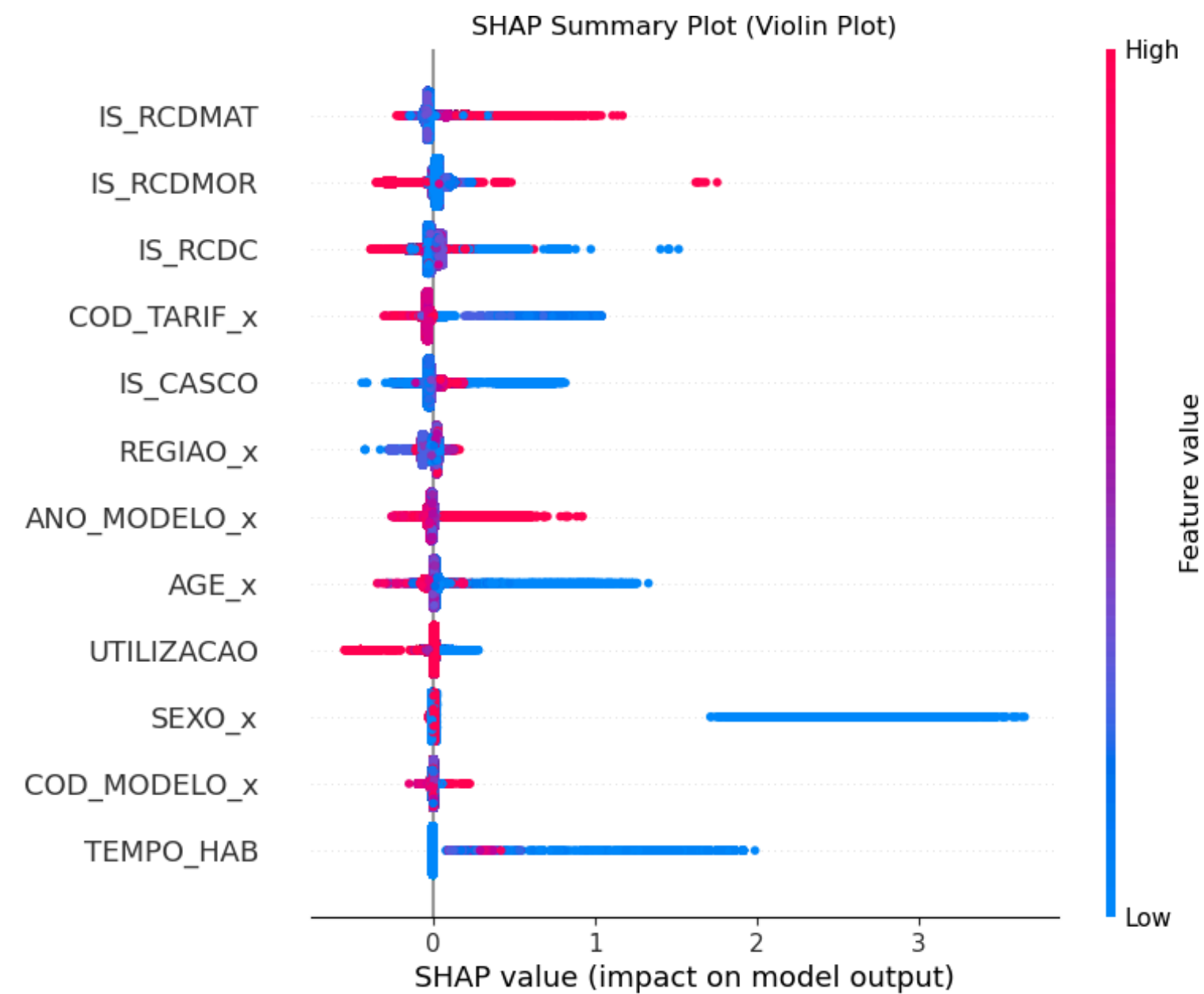
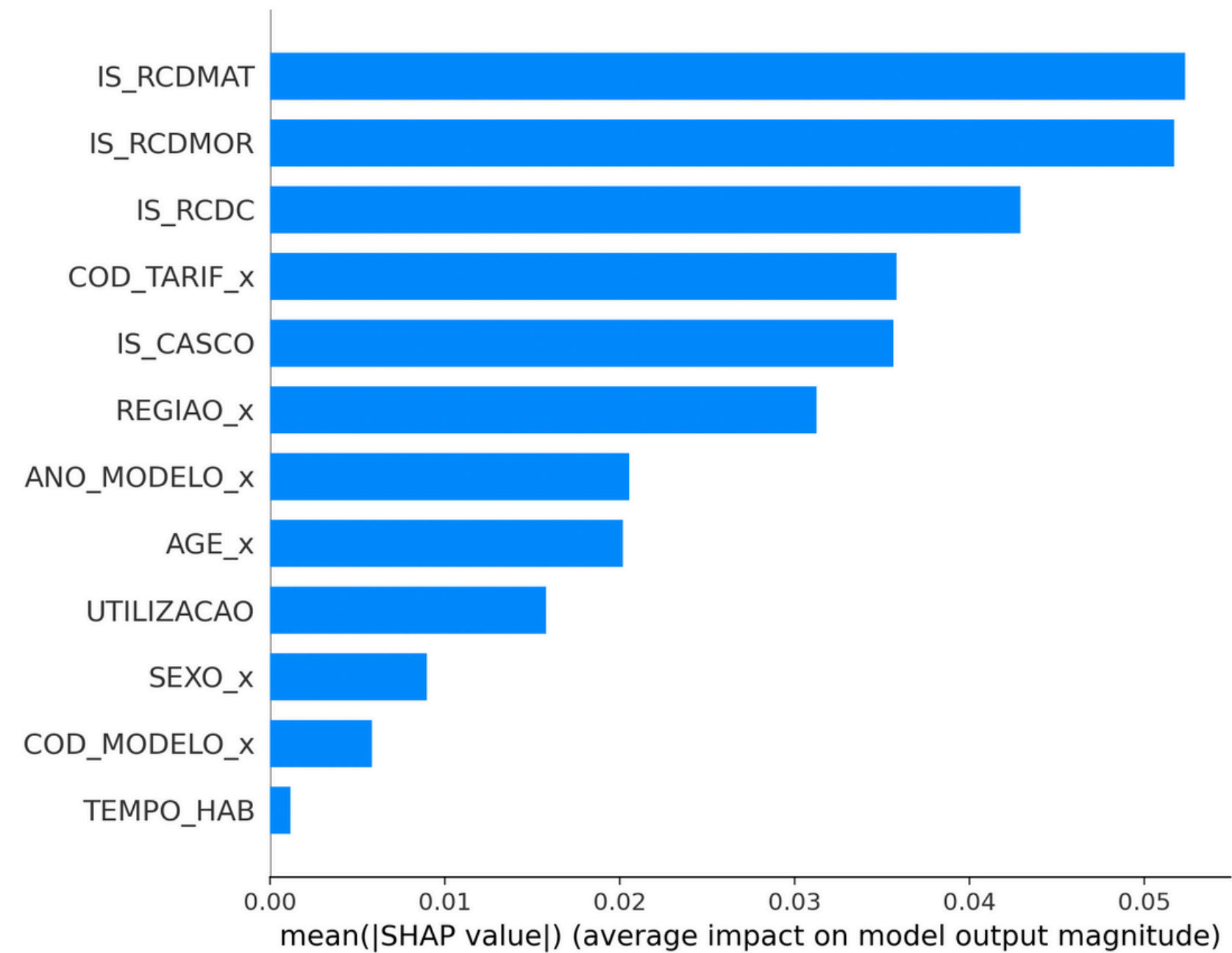
Etude de la fréquence : XGBoost

Séparation du jeu de données + **3-Fold Cross validation**



MAE	MSE	RMSE
1.0021	2.3483	1.5324

Importance des variables (Shap) XGBoost



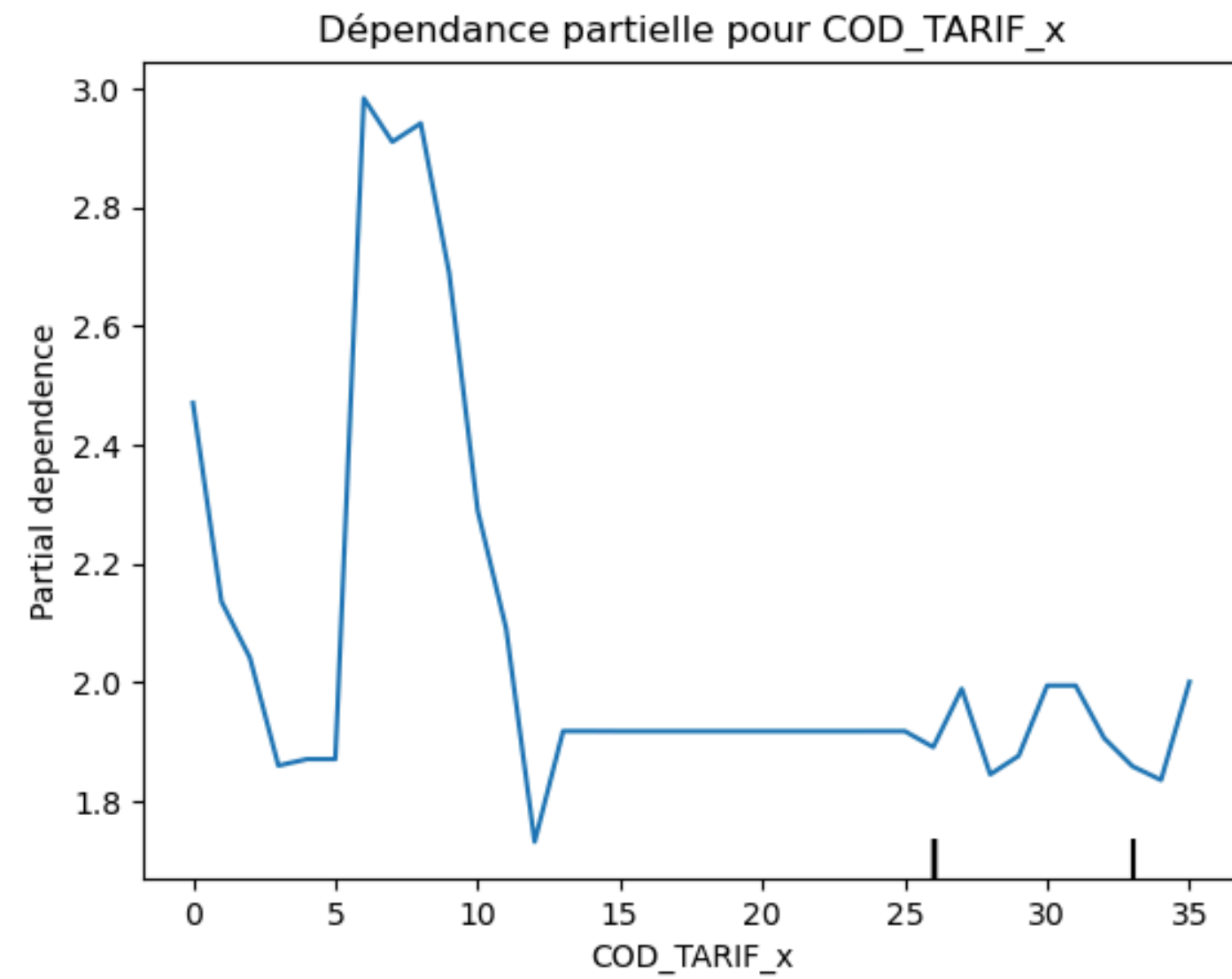
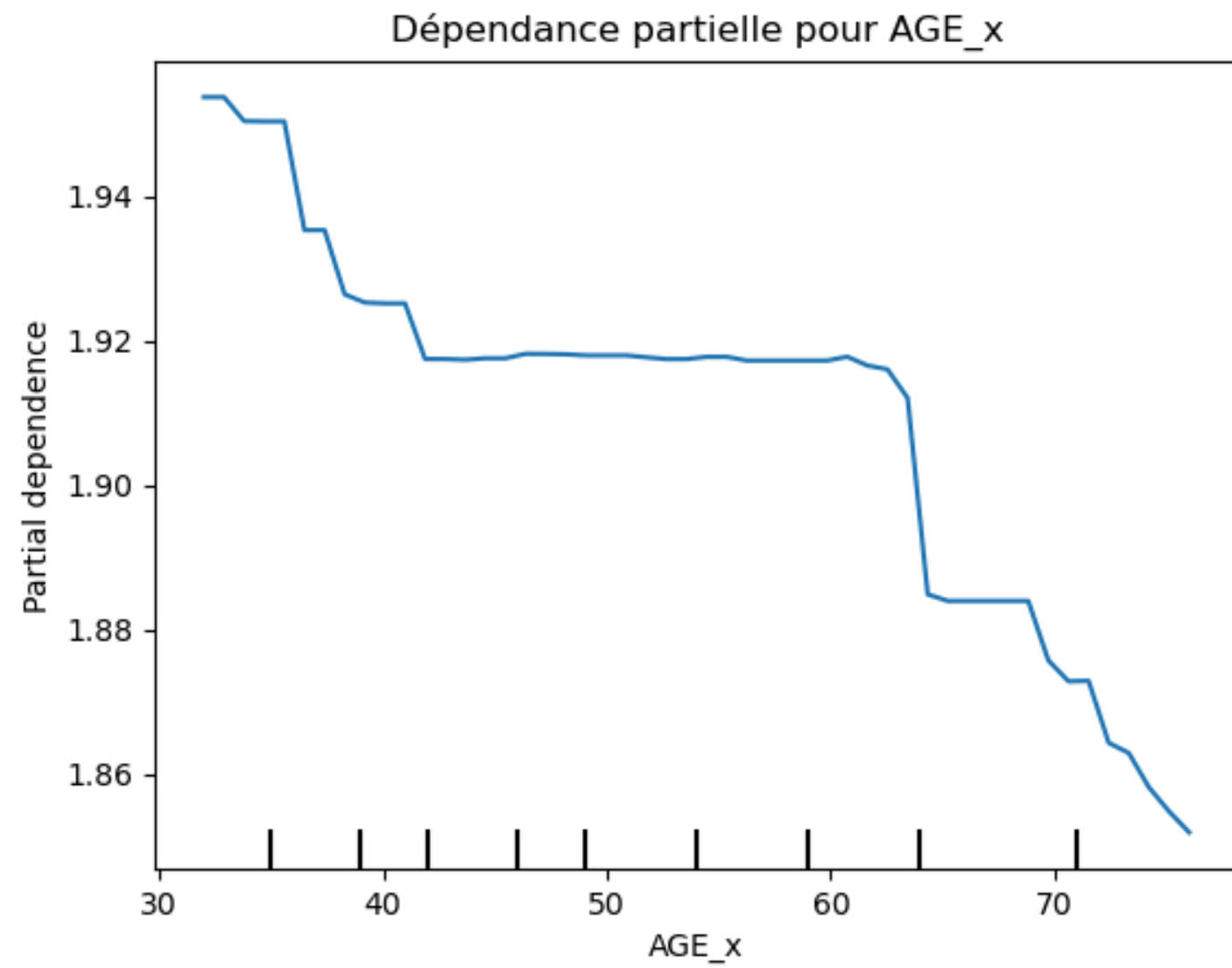
Etude de la fréquence : Comparaison GLM et XGBOOST

Séparation du jeu de données pour le Xgboost : 3-Fold Cross validation

Modèle	MAE	MSE	RMSE
GLM poisson	0.9742	2.1003	1.4492
Xgboost	1.0021	2.3483	1.5324



Partial Dependence plot



Estimation de la prime pure à l'aide de notre modèle

5 nouveaux assurés

Police Number	Age	Region	Type véhicule	Coût Predit	fréquence Predite	PrimePure
1	20-25	Brasília	Véhicules tourismes	4070.25	1.93	7855.58
2	30-35	Toledo-cascavel	Véhicules tourismes	1370.15	2.75	3767.91
3	35-40	BA - Bahia	Bicyclettes, motos	161.64	1.67	269.93
4	40-45	Ribeirão Preto	Pick up lourd	4156.01	0,3	1246.80
5	55-60	Met. Porto Alegre	Pick up lourd	1781.81	0,2	356.36



**Merci pour votre
écoute !**

