

Pour construire mon modèle de prédiction du churn, j'ai délibérément choisi de me concentrer uniquement sur trois variables numériques : **tenure**, **MonthlyCharges** et **TotalCharges**. Ces variables sont, selon moi, les plus représentatives du comportement du client et de son engagement vis-à-vis du fournisseur.

tenure : représente la durée de la relation client-fournisseur. Plus un client reste longtemps, plus il est fidèle, ce qui est directement lié à la probabilité de churn.

MonthlyCharges : reflète les dépenses mensuelles du client. Un montant élevé ou fluctuant peut influencer la décision de quitter le fournisseur.

TotalCharges : agrégation des dépenses depuis le début du contrat, ce qui donne une idée du niveau d'investissement du client.

Les autres variables ont été exclues pour les raisons suivantes :

Variables d'identification ou personnelles, comme **customerID** ou **gender**, n'apportent aucune valeur prédictive pertinente.

D'autres variables comme **Partner**, **Dependents**, ou encore **SeniorCitizen** influencent indirectement les charges, déjà prises en compte via **MonthlyCharges** et **TotalCharges**.

Les options de services (**PhoneService**, **InternetService**, **StreamingTV**, etc.) affectent déjà les coûts mensuels et sont donc redondantes avec les variables financières retenues.

Enfin, des variables comme **Contract** ou **PaymentMethod** traduisent des comportements qui peuvent être partiellement captés par **tenure** et les charges.