

Prédiction du Risque de Crédit et Segmentation Stratégique par le Machine Learning

Une approche data-driven pour optimiser le couple Risque/Profit

Présenté par : BELOUH Amine & BABA Abdellah

Sous la supervision de : Prof. BAKHER Zine Elabidine



Le Dilemme Central : L'Équilibre entre Risque de Perte et Opportunité de Croissance

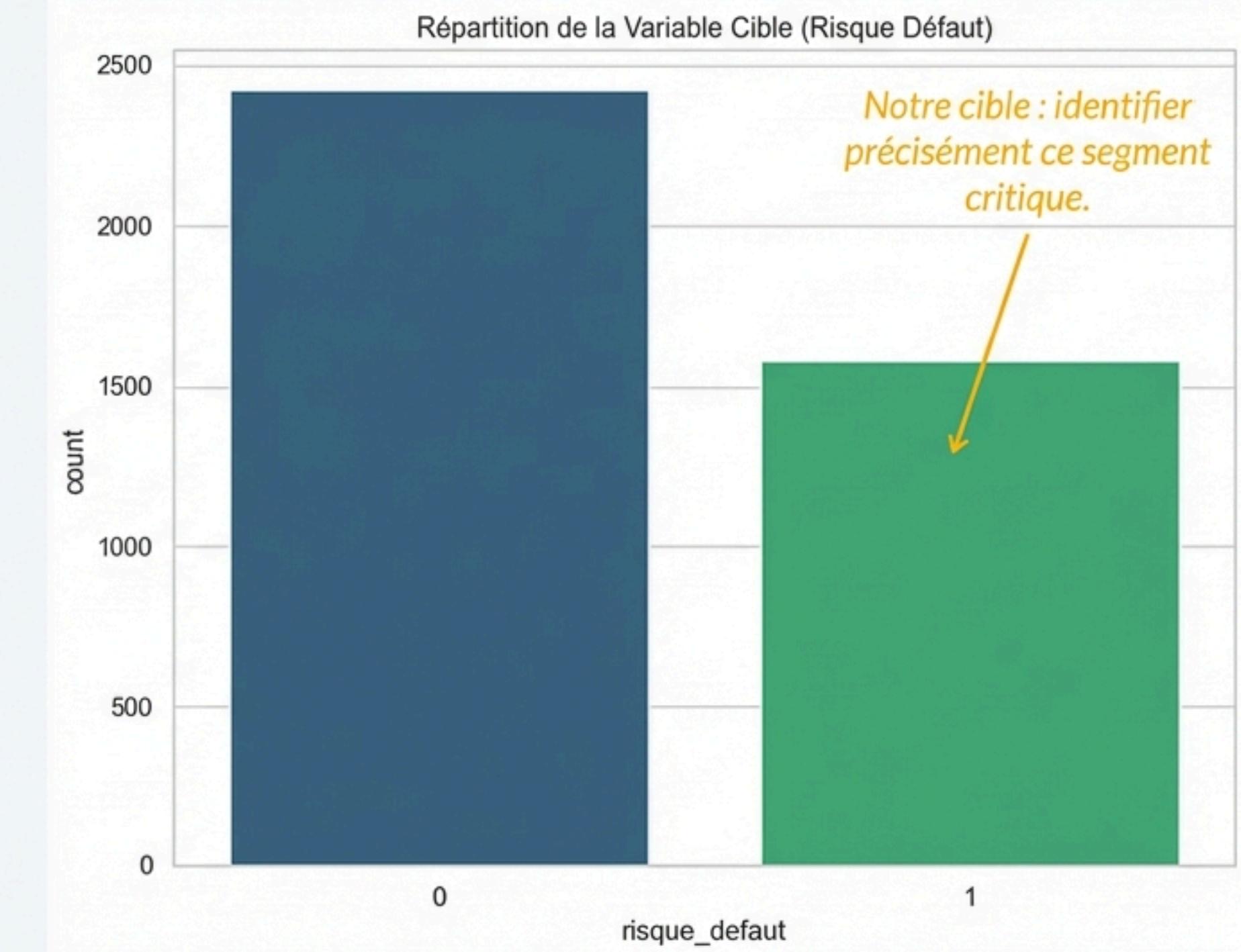


Objectif 1 : Minimiser le Risque.
Éviter de prêter à des clients qui
seront en défaut.



Objectif 2 : Maximiser le Profit.
Ne pas refuser de clients solvables
par excès de prudence.

Le Défi : Notre portefeuille actuel présente un
nombre significatif de défauts potentiels. La
question est de savoir comment les identifier
en amont avec une précision chirurgicale.



Notre Approche : Une Fondation de Données Robuste

1. Nettoyage & Inspection

Analyse d'un dataset de 4 000 clients.
Suppression des identifiants non prédictifs pour se concentrer sur les signaux pertinents.



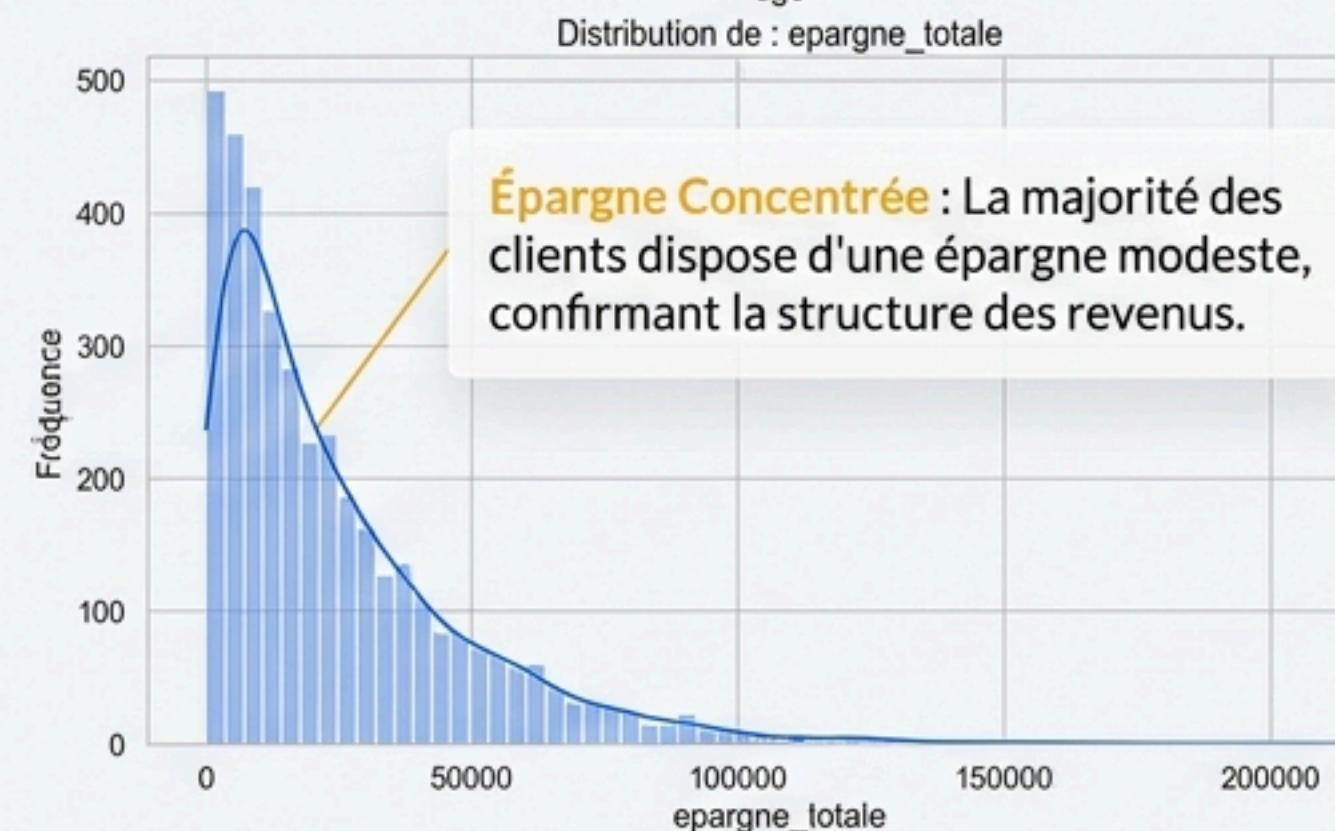
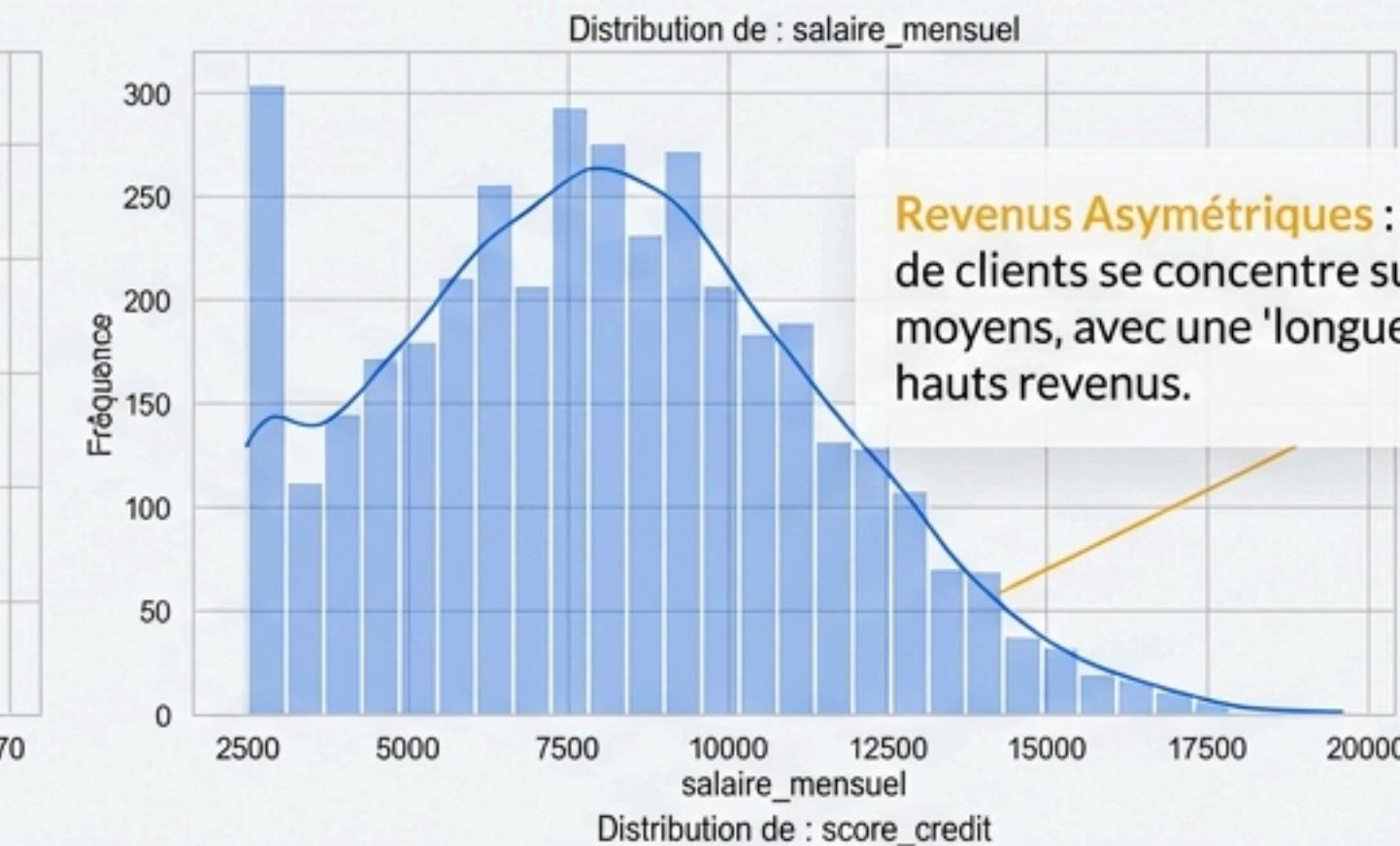
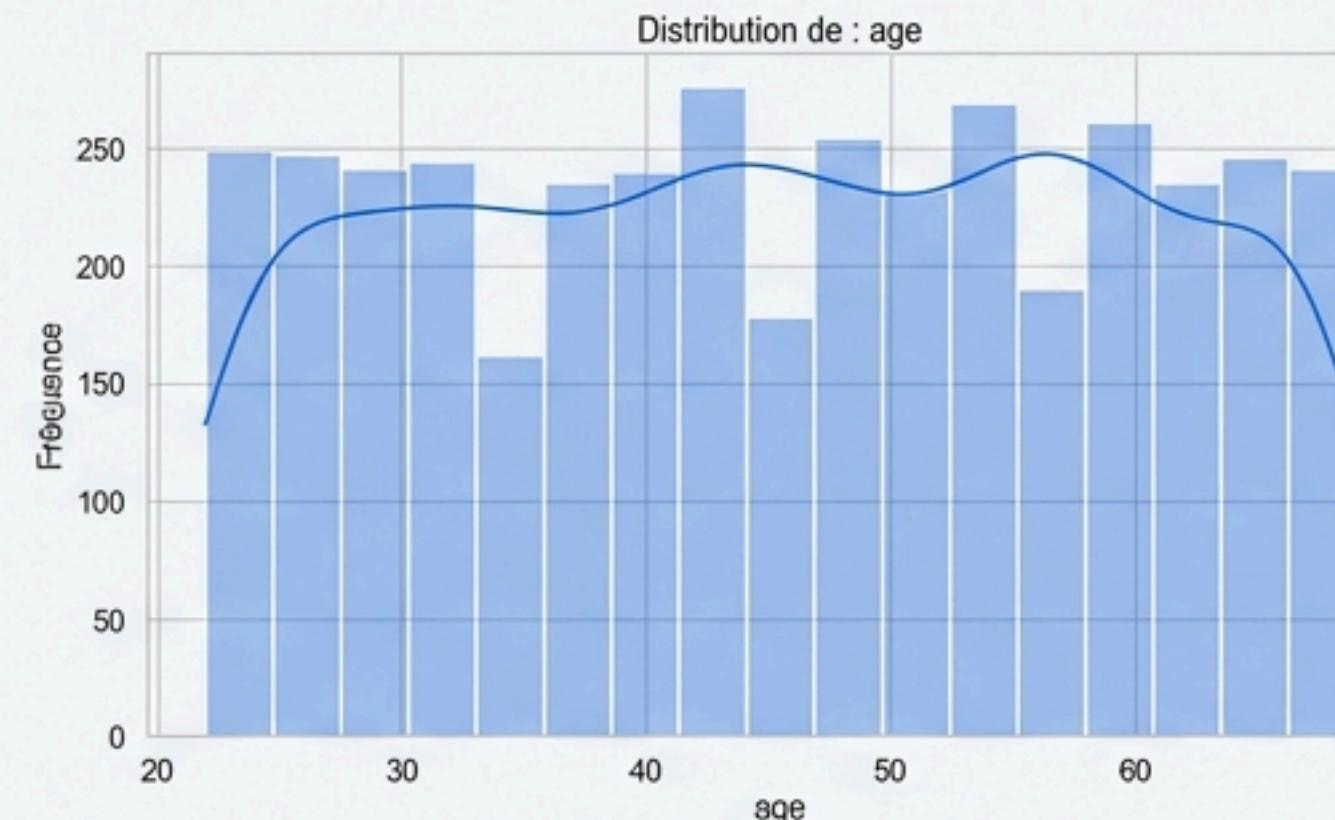
2. Imputation Stratégique

Traitement des données manquantes (`salaire_mensuel`, `epargne_totale`) via une **imputation par la médiane**. Ce choix neutralise l'impact des hauts revenus extrêmes, contrairement à la moyenne.

3. Standardisation & Encodage

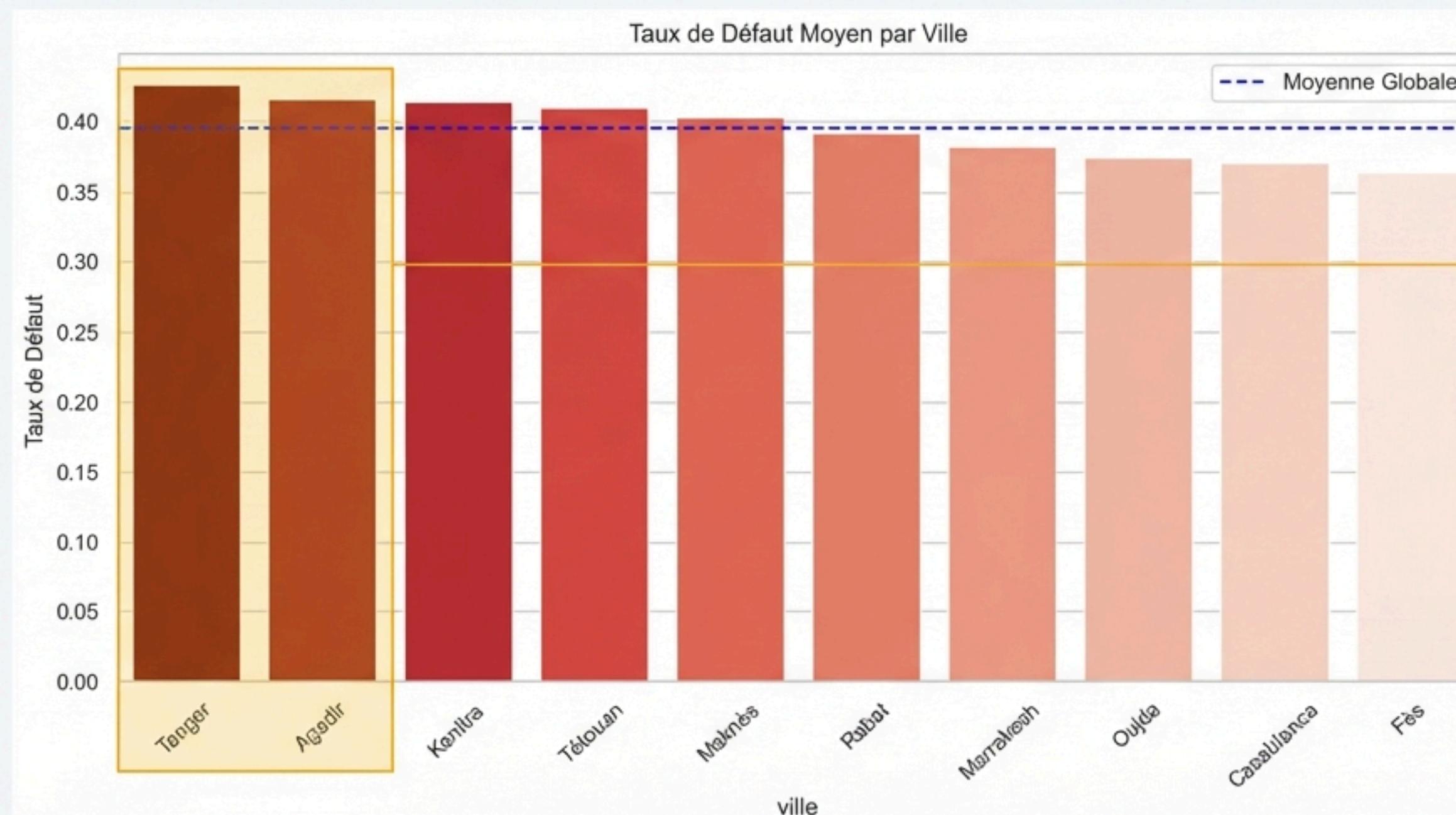
Préparation des variables (StandardScaler, LabelEncoder) pour garantir une interprétation équitable par les algorithmes.

Qui sont nos Clients ? Premiers Insights de l'Analyse Exploratoire



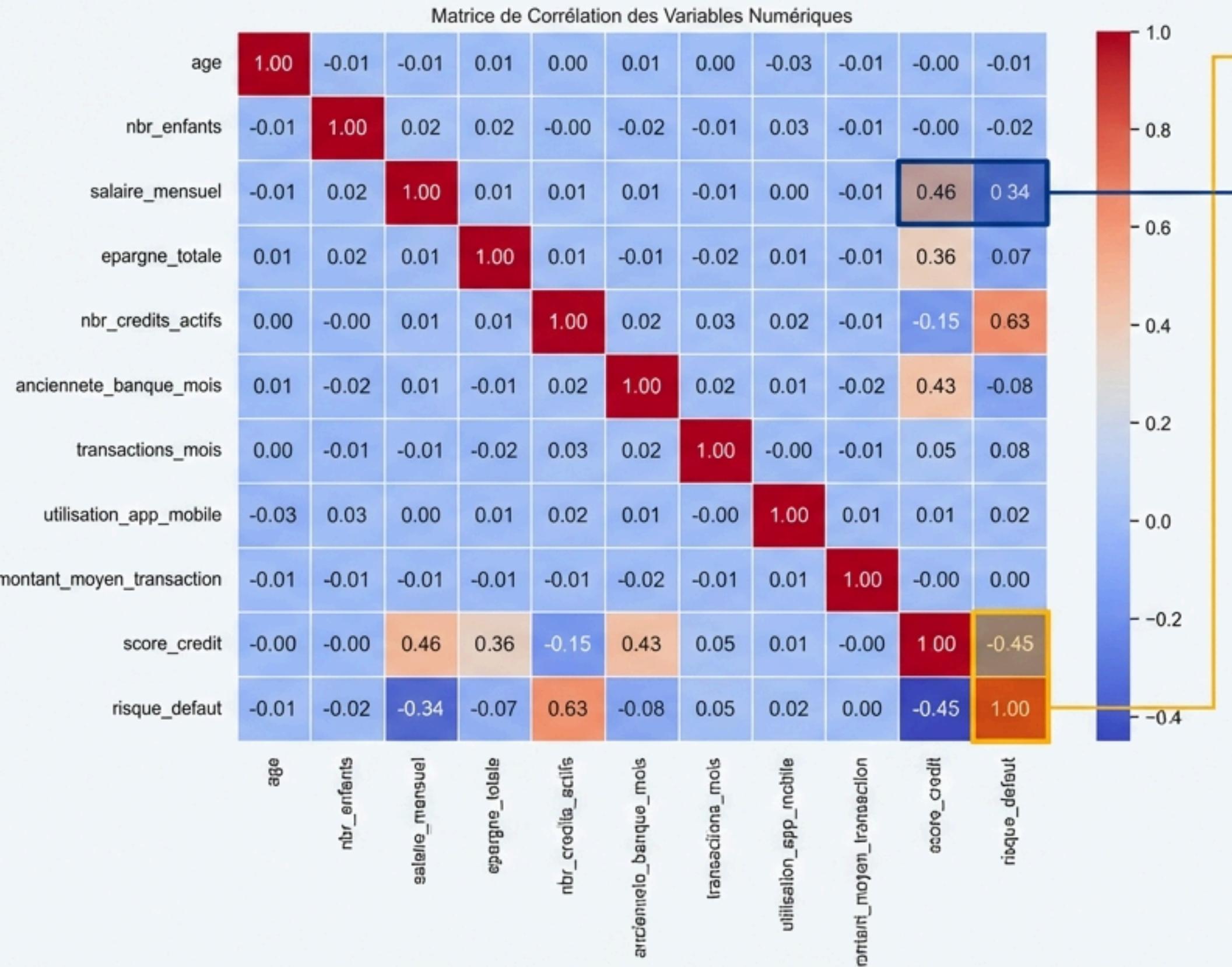
Analyse Géographique : Le Risque a une Adresse

Le taux de défaut n'est pas uniforme sur le territoire. Certaines villes présentent une surexposition significative au risque, nécessitant une vigilance accrue.



Tanger et Agadir : Taux de défaut significativement supérieurs à la moyenne.
Une analyse locale et un ajustement du risque s'imposent.

Au Cœur des Relations : Qu'est-ce qui Influence le Défault ?

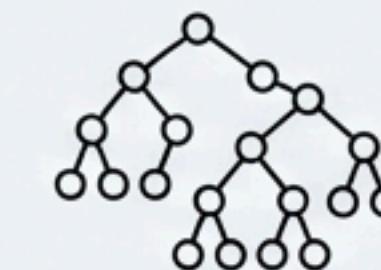
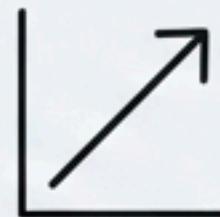


• **Lien Prédicatif Majeur :**
Corrélation négative forte de **-0.45**.
Plus le score de crédit baisse, plus le risque de défaut augmente. C'est notre levier de prédiction principal.

• **Logique Métier Confirmée :**
Corrélation positive de **0.46**. Des revenus plus élevés sont, logiquement, associés à une capacité d'épargne supérieure.

La Course à la Performance : Random Forest vs. Régression Logistique

Pour garantir une performance maximale, nous avons comparé un modèle de référence simple et interprétable à un algorithme d'ensemble puissant, optimisé de manière exhaustive.



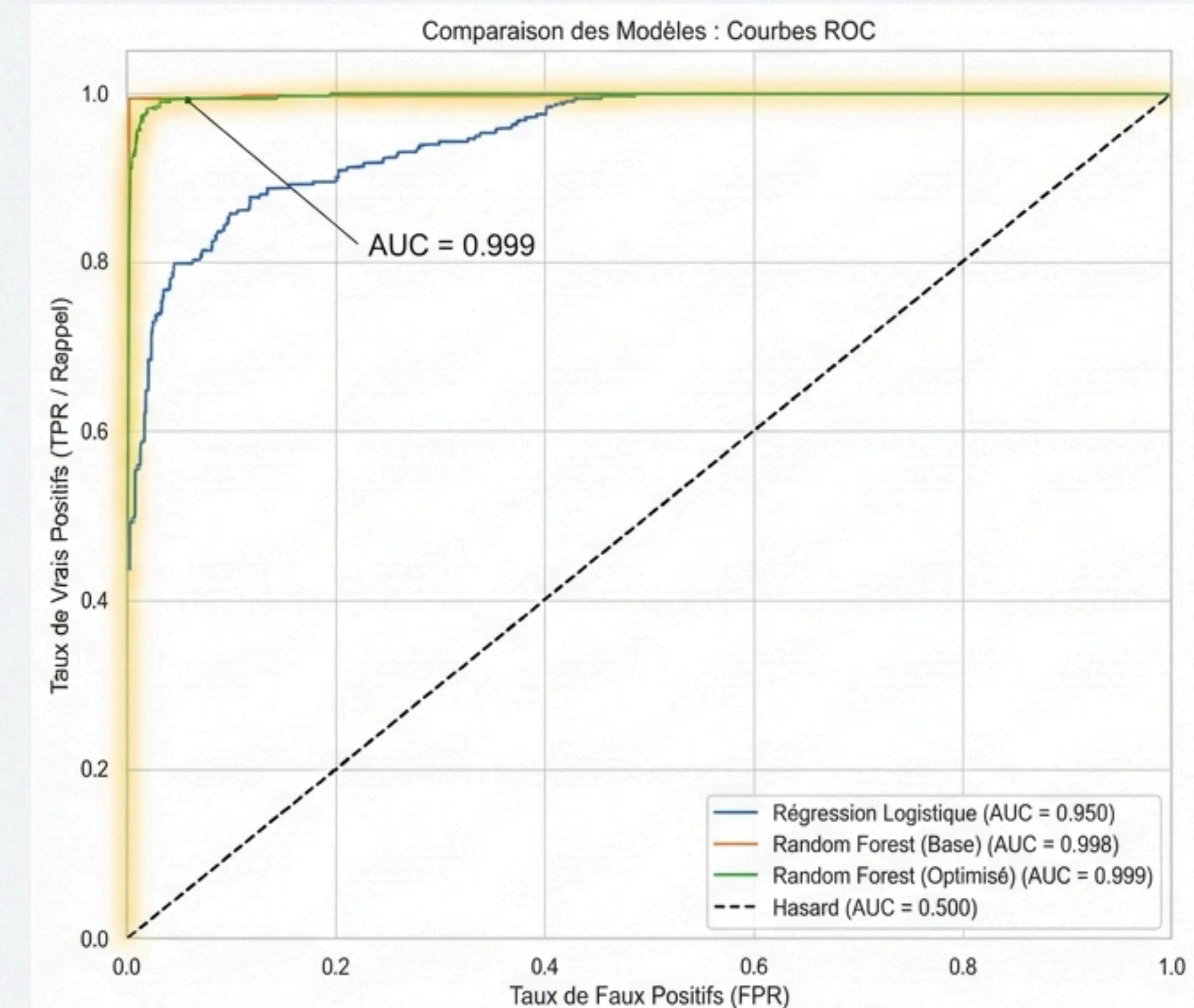
Caractéristique	Régression Logistique (Baseline)	Random Forest (Challenger)
Type de Modèle	Linéaire, Interprétable	Ensembliste, Non-linéaire
Principal Avantage	Simplicité, Rapidité	Puissance, Capture d'interactions complexes
Niveau d'Optimisation	Standard	GridSearchCV avec validation croisée (3-fold)

Verdict des Modèles : Une Précision Exceptionnelle

AUC = 0.999

(Random Forest Optimisé)

Notre modèle atteint un niveau de discernement proche de la perfection pour distinguer un client solvable d'un client à risque.

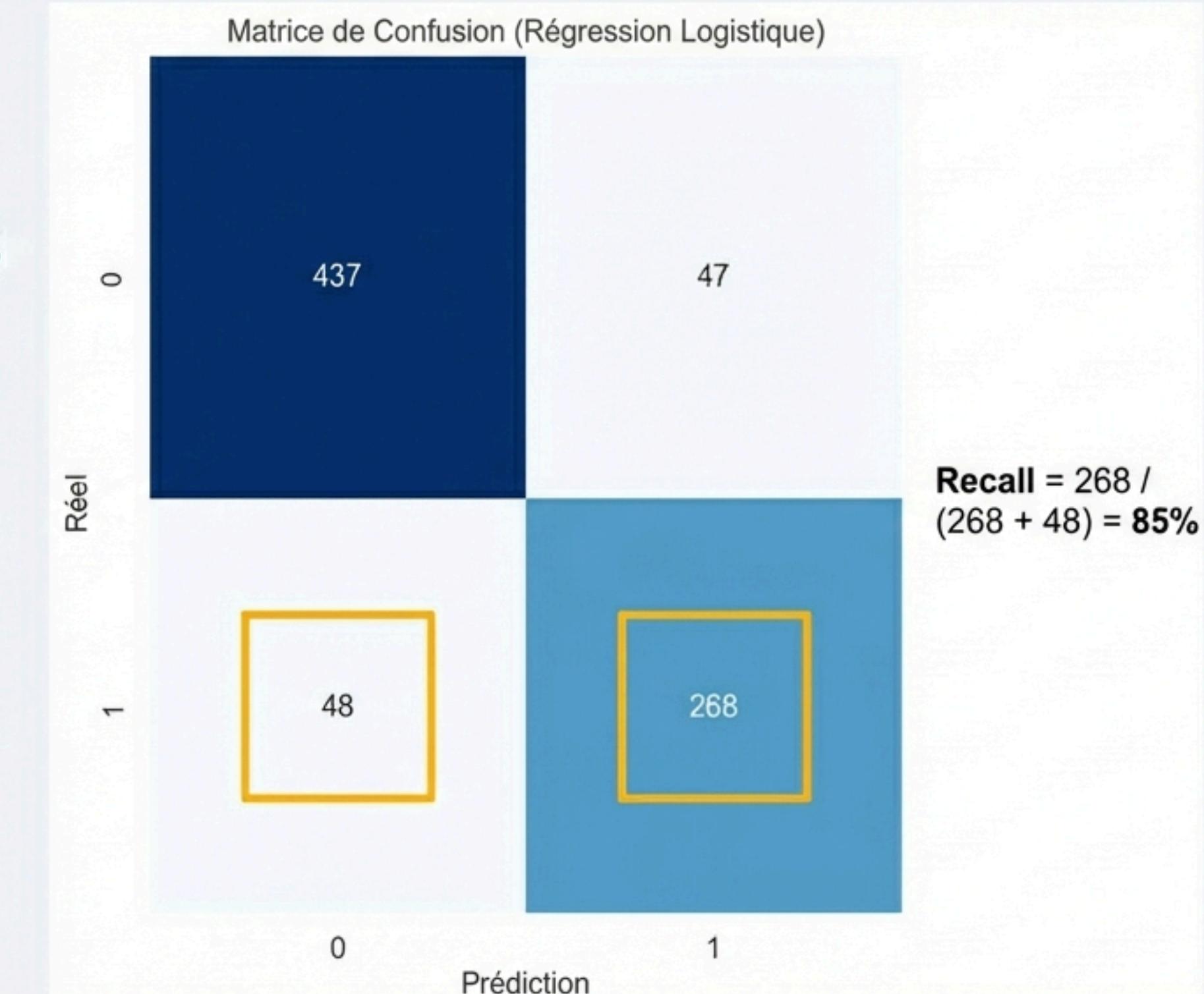


Au-delà du Score : L'Impact Direct sur la Minimisation des Pertes

85% de Recall

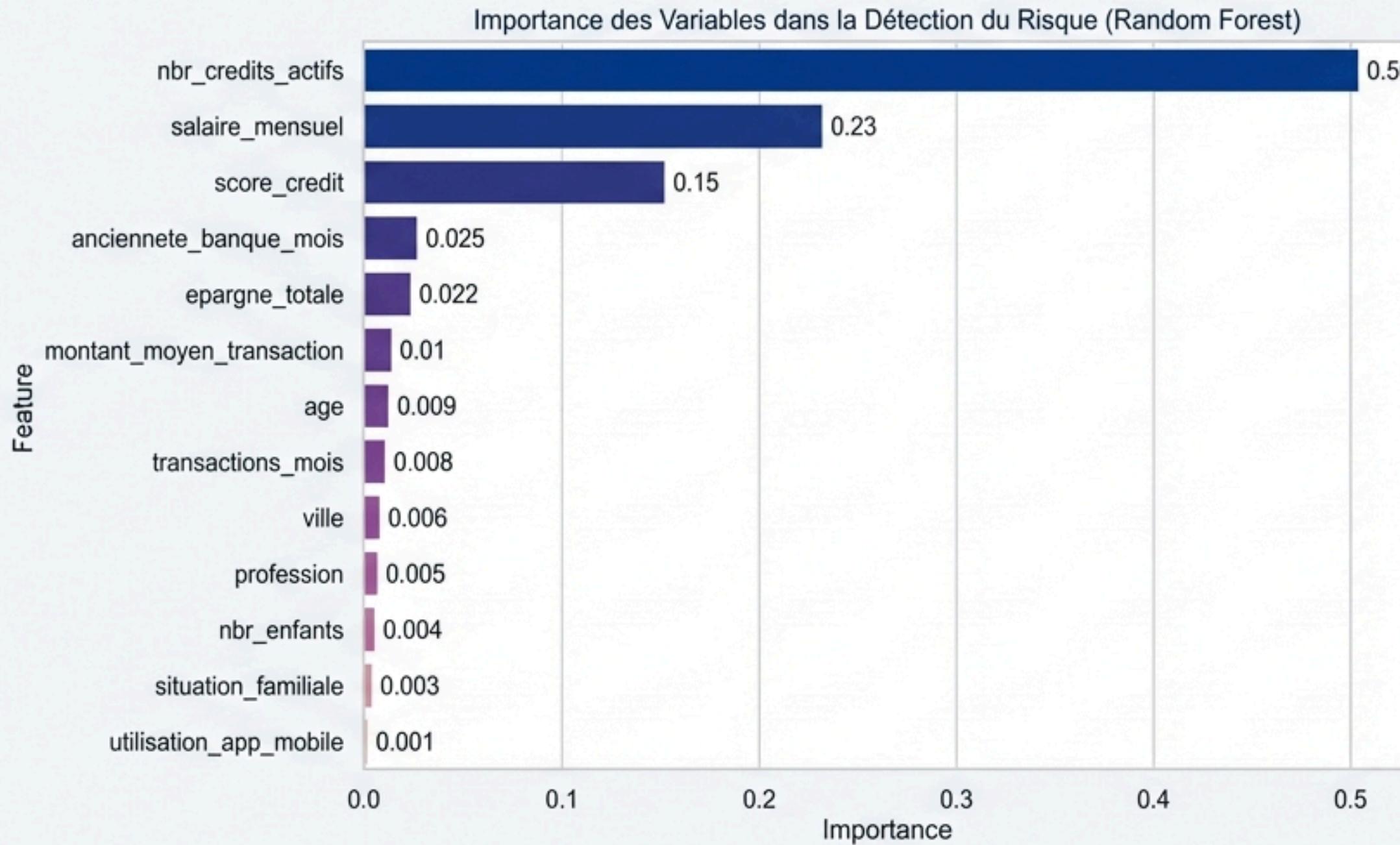
Interprétation business (Lato Regular)

Notre modèle identifie correctement **85% de tous les clients qui feront réellement défaut**. Sur 100 'mauvais payeurs', nous en signalons 85 en amont.



Le Moteur de la Décision : Sur Quoi se Base le Modèle ?

Le modèle n'est pas une boîte noire. Ses prédictions reposent sur des facteurs métier logiques et quantifiables. Le signal de risque le plus fort est le surendettement existant.



Top 3 Levier de Décision



N°1 - Nombre de Crédits Actifs ('nbr_credits_actifs') : Le facteur le plus déterminant.



N°2 - Salaire Mensuel : La capacité de remboursement.

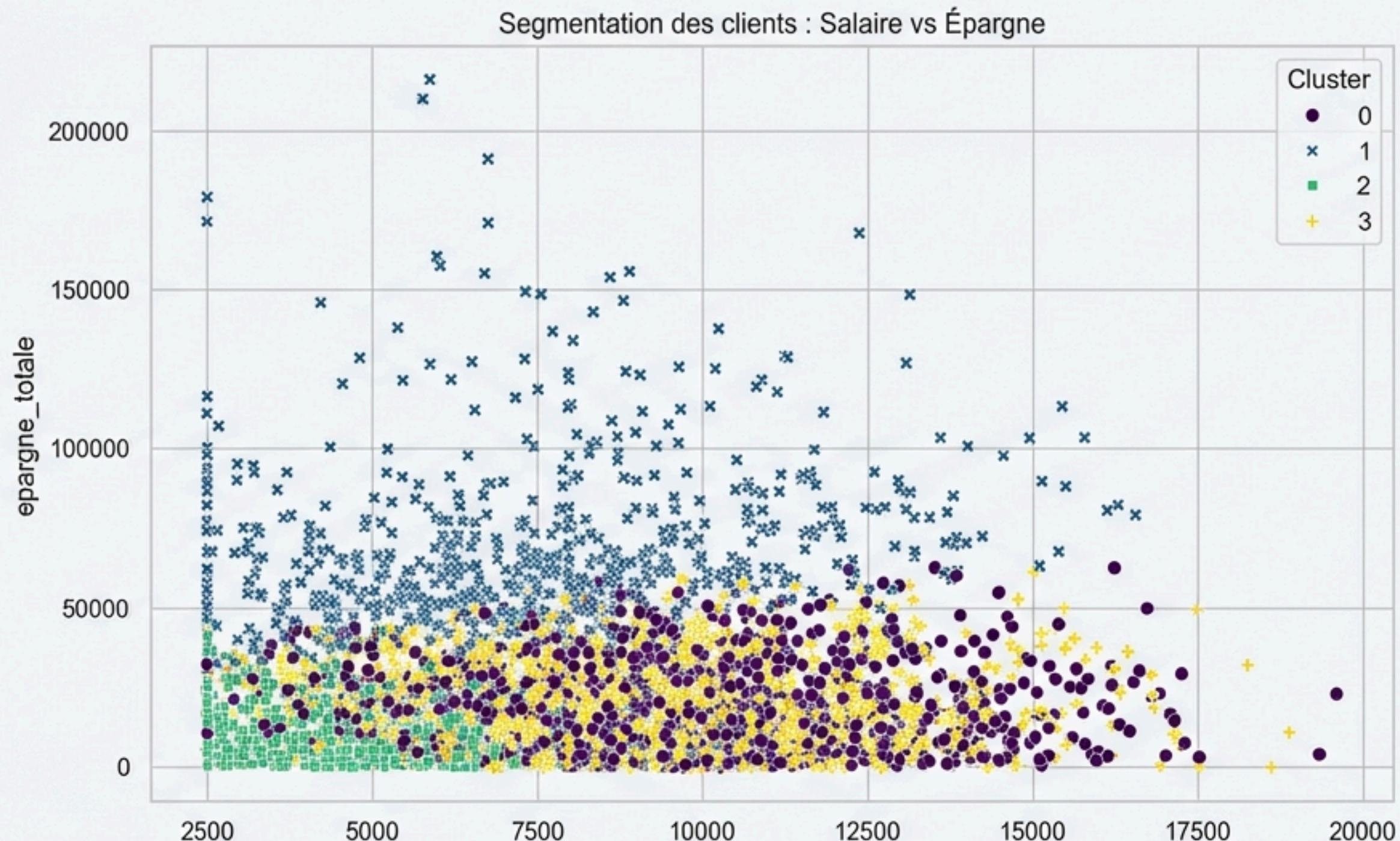
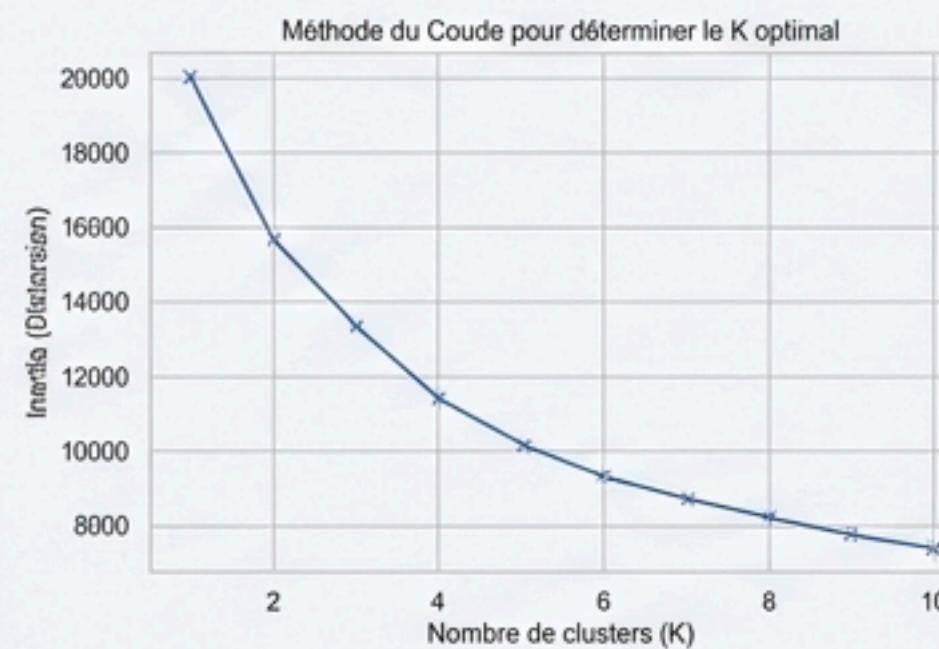


N°3 - Score de Crédit : L'historique de confiance.

De la Prédiction à la Personnalisation : 4 Profils Clients Stratégiques

Méthodologie

K-Means appliqué sur les données.
Le nombre optimal de 4 clusters a été validé par la méthode du coude.



- Cluster 1 (Premium) : Hauts revenus, épargne élevée. **Risque très faible.**
- Cluster 2 (Profils Fragiles) : Faible score, multiples crédits. **Risque élevé.**

- Cluster 3 (Familles Stables) : Revenus stables, forte ancienneté. **Risque faible.**
- Cluster 0 (Jeunes Actifs Digitalisés) : Salaires modérés, forte utilisation de l'app mobile. **Risque modéré.**

Vers une Banque Prédictive : Recommandations et Prochaines Étapes



Optimisation Opérationnelle

- **Approbation Automatisée** pour les profils 'Premium' (Cluster 1) avec un score de risque < 5%.
- **Alerte & Audit Manuel Systématique** pour tous les dossiers du segment 'Profils Fragiles' (Cluster 2).



Stratégie Commerciale Ciblée

- Lancer des campagnes de **cross-selling** (produits d'investissement) pour les 'Premium'.
- Développer des offres **d'accompagnement** et de micro-crédits pour les 'Jeunes Actifs Digitalisés'.



Gouvernance du Modèle

- Planifier un **ré-entraînement semestriel** pour s'adapter à l'évolution de l'économie.
- Mettre en place un **monitoring éthique** pour prévenir les biais (notamment géographiques).

Ce projet offre les clés pour passer d'une gestion de risque **réactive à une gestion proactive, personnalisée et pilotée par la donnée**.
