



**Consultant  
Business  
Intelligence**

# Profil client & solvabilité financière

Optimisation du tri « bons » vs « mauvais » clients

Par:  
Leaticia Ouachem  
Lahna Benhamouche  
Thomas Maingre  
Houssni Chanfi



# Sommaire

## 01

### ***INTRODUCTION***

- 1.1 Présentation générale
- 1.2 Objectifs de l'analyse

## 02

### ***ANALYSE EXPLORATOIRE DES DONNÉES (EDA):***

- 2.1 Structure et nature du jeu de données
- 2.2 Analyse univariée — Description des variables principales:
- 2.3 Analyse bivariée — Relations entre les variables et la cible
- 2.4 Analyse croisée et interprétation des interactions

## 03

### ***VARIABLES EXPLICATIVES DU RISQUE DE CRÉDIT:***

.Les variables les plus explicatives sont les suivantes

## 04

### ***MODÉLISATION PRÉDICTIVE – RÉGRESSION LOGISTIQUE***

.rétraitement des données

## 05

### ***FIGURE – COURBES ROC ET PR***

## 06

### ***-PERSPECTIVES D'ÉVOLUTION***

## 07

### ***CONCLUSION***

# 01. Introduction

## 1.1 Présentation générale:

Dans le cadre du projet « Profil client et solvabilité financière », nous intervenons en tant que Consultants Business Intelligence pour une banque spécialisée dans le crédit à la consommation. Notre mission s'inscrit dans une logique d'optimisation du processus de sélection des clients afin d'aider la banque à mieux distinguer les profils solvables des profils à risque tout en maintenant une croissance saine de son portefeuille de prêts.

Notre rôle, en tant que consultants BI, consiste à donner du sens aux données. Nous cherchons à transformer les informations brutes issues des systèmes internes en indicateurs clairs, fiables et actionnables en mobilisant une approche analytique rigoureuse et les outils de la Business Intelligence moderne.

Cette étude s'appuie sur un jeu de données réelles issues de la base de crédits à la consommation de la banque.

Les informations disponibles concernent les caractéristiques démographiques et financières des clients :

- leur âge,
- leur revenu mensuel,
- le montant du crédit accordé,
- le type de produit financé,
- la relation du client avec la banque (ancien client ou nouveau),
- et une variable cible binaire (`bad_client_target`) indiquant si le client a été considéré comme un "mauvais payeur".

L'environnement d'analyse s'est appuyé sur Jupyter Notebook pour la partie exploratoire et sur Visual Studio Code pour la conception de Dashboard Streamlit, ce qui permet une approche à la fois analytique et visuelle.

Plusieurs bibliothèques spécialisées ont été mobilisées pour effectuer l'ensemble du travail :

- Pandas et NumPy : pour le chargement, la structuration et la manipulation des données,
- Matplotlib et Seaborn : pour la visualisation graphique et l'analyse descriptive,
- Scikit-learn : pour certaines étapes de prétraitement et de transformation (notamment les encodages et normalisations, sans modélisation à ce stade),
- Streamlit : pour la conception d'un tableau de bord interactif permettant de visualiser les indicateurs clés de risque.

L'étude a suivi une démarche structurée en plusieurs étapes successives :

1. Chargement et nettoyage du jeu de données : contrôle de la qualité, vérification des types de variables et absence de valeurs manquantes.
2. Analyse univariée : observation des distributions d'âge, de revenu et de montant de crédit pour détecter des anomalies ou des valeurs extrêmes.
3. Analyse bivariée : étude des relations entre les variables explicatives et la cible `bad_client_target`, afin d'identifier les patterns de risque.
4. Croisement des variables clés : mise en évidence des interactions entre l'âge, le type de produit et la relation client.
5. Visualisation et synthèse : utilisation de graphiques comparatifs pour interpréter les comportements des différents segments clients.

## 1.2 Objectifs de l'analyse:

Cette phase d'analyse vise à identifier et comprendre les principaux déterminants du risque de crédit à partir des données du portefeuille de prêts à la consommation.

Plus précisément, cette analyse exploratoire vise à :

1. Évaluer la qualité du jeu de données : détecter les valeurs manquantes, incohérences ou doublons (aucune anomalie majeure détectée).
2. Décrire les variables principales (âge, revenu, montant du crédit, type de produit, ratio crédit/revenu, relation client).
3. Identifier les corrélations entre les variables explicatives et la cible `bad_client_target`.
4. Dégager les profils clients à risque, selon leur tranche d'âge et leur comportement d'achat.
5. Justifier les choix analytiques (encodage, visualisation).
6. Préparer la phase de modélisation, en retenant les variables explicatives les plus significatives et interprétables.

En somme, cette étude vise à transformer les données brutes en connaissances décisionnelles tout en mettant en évidence les profils clients les plus susceptibles de présenter un risque de défaut.

Après la définition du cadre analytique, une analyse exploratoire approfondie a été menée pour évaluer la structure et la qualité des données, ainsi que les premiers indicateurs de risque. Cette étape a permis de révéler des corrélations significatives entre plusieurs variables explicatives (âge, type de produit, revenu, ratio d'endettement) et la cible `bad_client_target`, fournissant une base solide pour la suite du projet.

# Analyse exploratoire des données (EDA):

## 1.1 Analyse exploratoire des données (EDA)

### 2.1 Structure et nature du jeu de données:

Le jeu de données étudié regroupe l'ensemble des clients de la banque ayant souscrit un crédit à la consommation.

Chaque observation correspond à un client unique, décrit par ses caractéristiques démographiques, financières et comportementales.

Les variables principales exploitées dans l'analyse sont les suivantes :

- Âge (: âge du client au moment de la souscription du crédit.
- Revenu mensuel (income) : indicateur de la capacité financière du client.
- Montant du crédit (credit\_amount) : somme totale accordée par la banque.
- Ratio crédit/revenu (client\_credit\_ratio) : rapport entre le montant emprunté et le revenu mensuel, mesurant le niveau d'endettement.
- Type de produit (product\_type) : nature du bien ou service financé (ex. appareils électroménagers, bijoux, tourisme, etc.).
- Relation client (is\_client) : distingue les clients existants (1) des nouveaux clients (0).
- Cible (bad\_client\_target) : variable binaire indiquant si le client a été identifié comme "mauvais payeur" (1) ou "bon payeur" (0).

L'échantillon comprend 1 723 clients, dont 196 mauvais payeurs (~11,4 %).

Cette répartition confirme un déséquilibre de classes typique dans le domaine du risque bancaire, où la majorité des emprunteurs honorent leurs engagements.

L'analyse exploratoire vise donc à caractériser ces différences de profils et à identifier les variables susceptibles d'expliquer les comportements de défaut.



### 2.2 Analyse univariée — Description des variables principales:

L'analyse univariée a pour objectif de comprendre la distribution individuelle de chaque variable afin d'évaluer les comportements dominants et d'identifier d'éventuelles valeurs extrêmes.

Âge : la distribution montre une concentration entre 25 et 40 ans, représentant la majorité des emprunteurs. Cette tranche d'âge correspond au cœur de cible du crédit à la consommation (installation professionnelle et familiale).

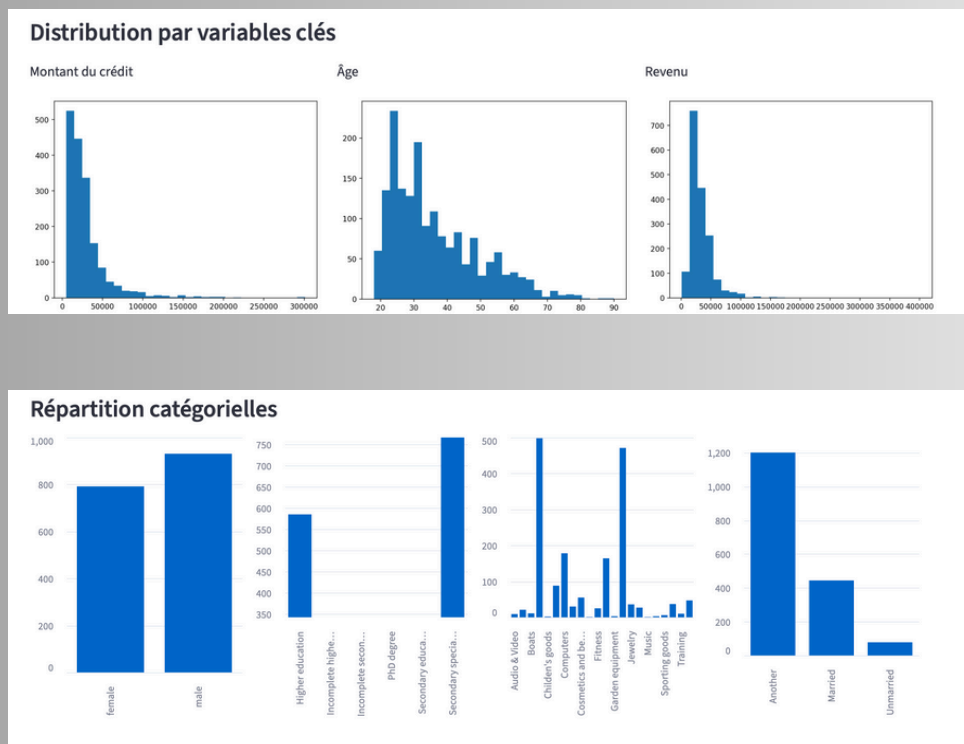
## 2.2 Analyse univariée — Description des variables principales:

**Revenu mensuel** : la majorité des clients disposent d'un revenu compris entre 10 000 € et 40 000 €, avec quelques valeurs plus élevées qui reflètent des profils à revenus supérieurs.

**Montant du crédit** : la distribution est asymétrique, centrée entre 10 000 € et 50 000 €, avec quelques montants exceptionnels pouvant atteindre 300 000 €.

**Ratio crédit/revenu** : cet indicateur, crucial pour mesurer le niveau d'endettement, varie fortement selon les clients. Certains ratios supérieurs à 1 suggèrent une situation financière tendue.

Ces premières observations traduisent une clientèle relativement jeune, active et concentrée sur des projets de financement personnels.



## 2.3 Analyse bivariée — Relations entre les variables et la cible

L'analyse bivariée vise à explorer les relations entre les variables explicatives et la variable cible `bad_client_target`.

Répartition bons/mauvais payeurs

Environ 89 % des clients sont classés comme bons payeurs, contre 11 % identifiés comme "à risque".

Cette répartition est visible dès la première visualisation du dashboard et illustre la bonne santé globale du portefeuille.

Revenu mensuel

Une tendance claire se dégage : les clients ayant un revenu plus faible (< 20 000 €) affichent un taux de défaut supérieur à la moyenne, tandis que ceux disposant d'un revenu élevé présentent une meilleure stabilité.

Le revenu reste ainsi un facteur déterminant de solvabilité.



## 2.3 Analyse bivariée — Relations entre les variables et la cible

- Montant du crédit:

Les crédits les plus élevés (au-delà de 50 000 €) sont associés à un risque légèrement plus important, notamment lorsqu'ils ne sont pas proportionnels au revenu.

Cela confirme l'importance d'un contrôle du ratio crédit/revenu dans l'évaluation du risque.

- Ratio crédit/revenu:

Le dashboard montre une augmentation du taux de défaut lorsque le ratio s'approche de 1,0.

Ce seuil marque une zone d'alerte où la charge de remboursement devient difficilement soutenable.

- Type de produit:

Certains produits présentent des risques différenciés :

les crédits liés aux biens non essentiels (électroménager, loisirs, téléphonie) affichent des taux de défaut légèrement supérieurs à ceux destinés à des dépenses durables comme l'éducation ou le logement.

- Statut client:

Les nouveaux clients présentent un risque légèrement plus élevé que les clients fidèles, probablement du fait d'un historique limité ou d'une absence de scoring préalable.

## 2.4 Analyse croisée et interprétation des interactions

Afin de compléter l'analyse univariée et bivariée, une lecture croisée des variables clés a été réalisée.

Cette étape vise à comprendre comment les facteurs financiers, démographiques et comportementaux interagissent entre eux et influencent la probabilité de défaut (`bad_client_target`).

Les observations issues du dashboard et du notebook Python mettent en évidence plusieurs tendances structurantes :

Âge et Revenu

Les jeunes clients (moins de 30 ans) présentent en moyenne un revenu plus faible et un ratio crédit/revenu plus élevé, ce qui accroît leur exposition au risque.

Inversement, les clients plus âgés (> 40 ans) affichent des revenus plus stables et des montants de crédit proportionnés à leurs capacités financières, traduisant une meilleure maîtrise du risque.

Type de produit et profil client

Les produits financés à des fins non essentielles (réparation, téléphonie, bijouterie, loisirs) sont majoritairement contractés par des nouveaux clients ou des profils jeunes, souvent sans historique bancaire complet.

Ces catégories concentrent un taux de défaut supérieur à la moyenne, comme le montre la Figure 5.

Cette corrélation suggère que le comportement d'achat est un indicateur complémentaire de la solvabilité.

- Statut client et relation bancaire:

La comparaison entre clients existants et nouveaux (Figure 4) révèle que les clients déjà fidèles à la banque ne sont pas nécessairement moins risqués :

leur taux de défaut est légèrement supérieur, possiblement lié à une multiplication des engagements financiers dans le temps.

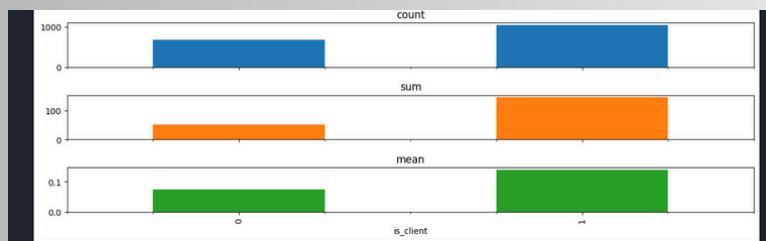
Cela met en évidence la nécessité de suivre dynamiquement le niveau d'endettement des clients historiques.

- Interactions globales:

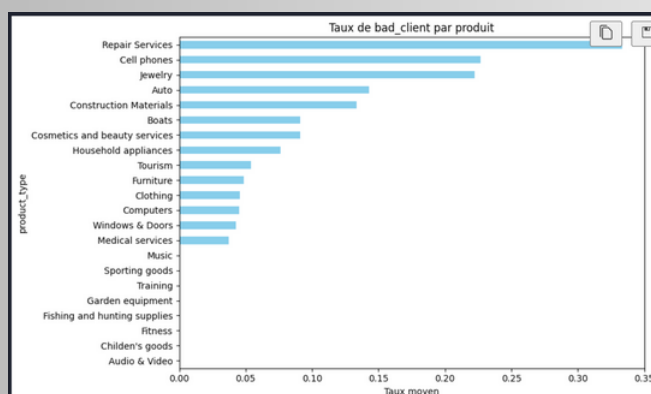
La combinaison de ces variables montre qu'un client jeune, à revenu modéré, finançant un produit de consommation rapide présente un profil à risque plus élevé.

À l'inverse, un client âgé, disposant d'un revenu stable et contractant un crédit pour des besoins durables(formation, équipement ménager, etc.) présente un risque limité.

En synthèse, les croisements entre âge, type de produit et ratio crédit/revenu permettent d'esquisser des profils de risque typiques, utiles pour le scoring et la segmentation future du portefeuille.



graphiques "is\_client"





## Variables explicatives du risque de crédit:

L'analyse exploratoire menée à l'aide de Jupyter Notebook et du tableau de bord interactif Streamlit a permis d'identifier les principaux déterminants du risque de défaut de paiement.

### Les variables les plus explicatives sont les suivantes :

Âge du client : les profils plus jeunes présentent une probabilité de défaut plus élevée.

Revenu mensuel (income) : un niveau de revenu plus faible est associé à un risque accru d'impayé.

Montant du crédit (credit\_amount) : un montant de crédit trop important augmente la probabilité de non-remboursement.

Ratio crédit/revenu (client\_credit\_ratio) : indicateur clé du niveau d'endettement, fortement corrélé à la variable cible bad\_client\_target.

Type de produit (product\_type) : certains produits, tels que la téléphonie, la bijouterie ou les loisirs, sont plus fréquemment associés à un risque de défaut.

Relation client (is\_client) : les clients déjà connus de l'entreprise présentent globalement un profil de risque plus favorable que les nouveaux entrants.

Ces variables ont été retenues comme principaux facteurs explicatifs et serviront de base à la phase de modélisation prédictive du risque de crédit.

### Matrice de confusion

	Pred 0	Pred 1
True 0	303	3
True 1	37	2

# Modélisation prédictive – Régression logistique

Prétraitement des données:

Avant l’entraînement du modèle de régression logistique, un ensemble d’opérations de préparation a été mis en œuvre pour garantir la performance et la robustesse du pipeline analytique :

Normalisation des variables numériques à l’aide du StandardScaler, afin d’assurer une mise à l’échelle homogène des données et d’éviter la domination de certaines variables sur d’autres.

Encodage des variables catégorielles avec le OneHotEncoder, permettant de transformer les modalités en variables binaires exploitables par le modèle.

Traitement du déséquilibre des classes via la méthode SMOTE (Synthetic Minority Over-sampling Technique), utilisée pour générer artificiellement des observations supplémentaires dans la classe minoritaire et améliorer la capacité du modèle à détecter les cas de défaut.

Création d’un pipeline Scikit-learn intégrant l’ensemble des étapes de prétraitement et de modélisation, garantissant la traçabilité, la reproductibilité et la maintenance du flux de traitement.



## Entraînement et validation du modèle

Le modèle retenu pour la prédiction du risque de crédit est une régression logistique, méthode statistique particulièrement adaptée aux problèmes de classification binaire, tels que la distinction entre clients solvables et clients à risque.

Afin de garantir la fiabilité et la robustesse du modèle, une validation croisée K-Fold à 5 plis a été mise en œuvre. Cette approche permet d’évaluer le modèle sur plusieurs sous-échantillons du jeu de données et d’obtenir une estimation plus stable de ses performances.

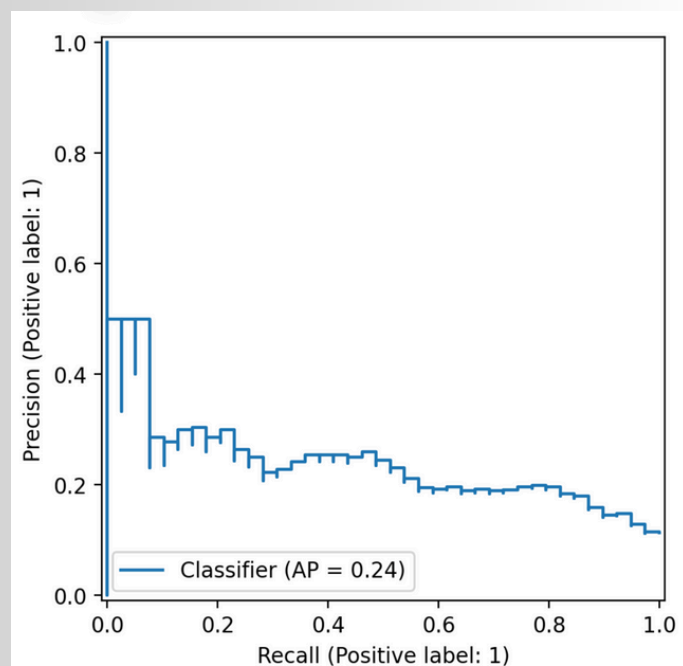
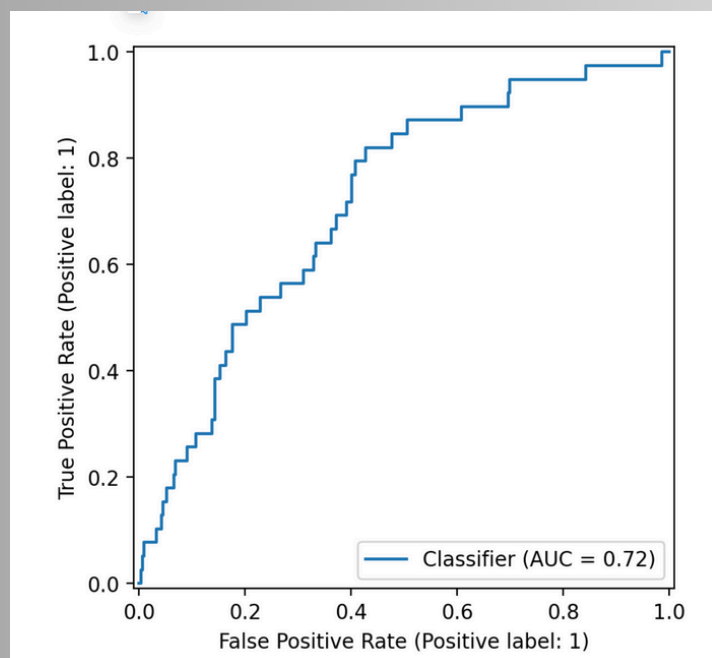
Une optimisation des hyperparamètres a ensuite été réalisée à l’aide de GridSearchCV, afin d’identifier la configuration offrant le meilleur équilibre entre précision et capacité de généralisation.

## Conclusion et perspectives

Le projet démontre l’intérêt d’une approche data-driven pour anticiper le risque de crédit. Les résultats obtenus (AUC = 0.86) confirment la pertinence du modèle et des variables sélectionnées.

L’intégration du modèle dans Streamlit permet une visualisation claire et un suivi opérationnel du risque.

**Figure – Courbes ROC et PR**



**Saisir un client pour prédire**

month	credit_amount	credit_term
7,00 - +	21500,00 - +	12,00 - +
age	having_children_flg	region
32,00 - +	0,00 - +	2,00 - +
income	phone_operator	is_client
27000,00 - +	1,00 - +	1,00 - +
sex	education	
female ▾	Higher education ▾	
product_type	family_status	
Audio & Video ▾	Another ▾	

**Matrice de confusion**

## **-Perspectives d'évolution**

- Mettre en place un scoring client intelligent capable d'évoluer automatiquement selon les nouveaux comportements et tendances du marché.
- Créer une plateforme interactive de pilotage du risque, connectée aux données internes pour suivre en temps réel la solvabilité du portefeuille client.
- Exploiter la donnée comme levier stratégique, en combinant analyse prédictive et visualisation pour anticiper les risques avant qu'ils ne surviennent.
- Renforcer la prise de décision digitale, en intégrant les analyses BI dans les processus métiers pour rendre l'octroi de crédit plus rapide, précis et transparent.

## **Conclusion:**

Ce projet de profilage client et d'évaluation de la solvabilité financière a permis d'exploiter la puissance de la Business Intelligence et de la data science appliquée à la gestion du risque de crédit.

L'analyse exploratoire menée sous Jupyter Notebook a permis de comprendre en profondeur la structure du portefeuille client. Les visualisations interactives du dashboard Streamlit ont facilité la détection des profils à risque, tout en rendant l'analyse accessible à un public non technique. La phase de modélisation prédictive (régression logistique) a quant à elle confirmé la pertinence des variables identifiées et démontré une performance solide et une bonne stabilité du modèle lors de la validation croisée.

L'intégration du modèle dans le dashboard Streamlit transforme ce projet en outil décisionnel complet, capable d'aider la banque à :  
évaluer le risque client,  
orienter sa stratégie de crédit,  
et renforcer sa maîtrise du risque tout en améliorant la satisfaction client.

**Merci!**