IUT de Paris - Rives de Seine
Université Paris Cité

Analyse et qualité des données d'un portefeuille de crédits automobiles

Rapport de qualité des données & Tableau de bord – Date de rendu : 01 décembre 2025

Bilal KHAFIFI-AZZAOUI
Abinaya SUBRAMANIAM
John-Etan UZAN

Enseignant : Jean-Philippe LAMANCHE

Introduction

L'analyse du risque de crédit occupe une place centrale dans les activités des établissements financiers, en particulier dans le contexte du financement automobile, un secteur caractérisé par un volume élevé de demandes et une diversité importante de profils emprunteurs. Dans ce cadre, la capacité à comprendre la structure d'un portefeuille de crédits, à évaluer la qualité des emprunteurs et à identifier les facteurs influençant l'octroi d'un financement constitue un enjeu stratégique majeur.

Le présent travail s'inscrit dans cette perspective. Il vise à analyser une base de données composée de 102 443 demandes de financement de véhicules, couvrant plusieurs années récentes. Cette base regroupe des informations variées relatives au véhicule financé, aux caractéristiques du prêt, au profil socio-économique du client et à la décision finale émise par la banque et le client. Avant d'envisager la construction d'outils d'aide à la décision (tels qu'un tableau de bord dynamique ou des modèles prédictifs), il est indispensable de procéder à une analyse descriptive approfondie et à une évaluation rigoureuse de la qualité des données.

L'objectif de cette première étape est double. D'une part, il s'agit de caractériser statistiquement les demandes de crédit afin de dégager les principales tendances du portefeuille : nature des véhicules financés, répartition des types de prêts, profils des emprunteurs, niveaux d'endettement, taux d'acceptation ou de refus, etc. D'autre part, l'étude vise à identifier les incohérences, valeurs aberrantes ou manquantes, susceptibles d'affecter la pertinence des analyses futures. Un premier nettoyage de la base est donc réalisé, permettant d'obtenir un jeu de données exploitable et cohérent.

1. Description des données

La base de données étudiée est constituée de **102 443 observations** et de **18 variables**, représentant des demandes de financement automobile. Les informations contenues dans cette base couvrent quatre grandes catégories : les caractéristiques du véhicule, les caractéristiques du crédit, les informations client, ainsi que les décisions prises par la banque et le client.

Avant toute analyse statistique, il est essentiel de présenter précisément la structure du jeu de données, afin de clarifier la nature et le rôle de chaque variable.

1.1. Structure générale de la base

La base se présente sous la forme d'un tableau où chaque ligne correspond à une demande de crédit, et chaque colonne à une variable décrivant cette demande. Toutes les variables sont dites observées, aucune variable dérivée ni variable cible (du type défaut/non défaut) n'est fournie dans l'état actuel.

Les observations couvrent plusieurs années récentes (2023–2025), permettant une analyse temporelle de l'activité.

1.2. Présentation des variables

Le tableau suivant synthétise les 18 variables de la base, en indiquant leur type, une description concise et leur domaine de valeurs observées.

Nom de la variable	Type	Description	Domaine observé
objet_vehicule	Catégorielle	Type d'usage du véhicule financé	VP, VU
classe_de_score	Catégorielle	Classe de risque attribuée au client	01 à 09
taux_interet	Numérique	Taux nominal du prêt en %	0 à 11,533
prix_achat	Numérique	Prix d'achat du véhicule (€)	1 500 à 128 500
montant_pret	Numérique	Montant du financement demandé (€)	1 400 à 109 346
duree_pret	Numérique	Durée du crédit (mois)	12 à 74
type_vehicule	Catégorielle	Nature du véhicule	VN, VO
type_pret	Catégorielle	Type de financement	LLD, LOA, VAC
taux_endettement	Numérique	Taux d'endettement du client (%)	0.29 à 99.9
code_postal	Catégorielle	Zone géographique du client	01000 à 98800 (codes métropole/DOM)
annee_demande	Numérique	Année de la demande	2023 à 2025
mois_demande	Numérique	Mois de la demande	1 à 20 (valeurs > 12 anormales)
jour_demande	Numérique	Jour de la demande	1 à 31
annee_naissance	Numérique	Année de naissance du client	1926 à 2006
mois_naissance	Numérique	Mois de naissance	1 à 12
jour_naissance	Numérique	Jour de naissance	0 à 31 (0 = anomalie)
etat_demande	Catégorielle	Décision de la banque	Octroyé, Refusé
decision_client	Catégorielle	Décision du client	Accord, Refus, Sans suite

2. Qualité des données

L'évaluation de la qualité des données constitue une étape essentielle avant toute analyse statistique. Elle permet d'identifier les incohérences, anomalies et valeurs manquantes susceptibles d'affecter la robustesse des résultats. Dans le cadre de ce projet, une attention particulière a été portée aux variables relatives au profil du client et aux caractéristiques du prêt, car elles jouent un rôle clé dans l'analyse du risque de crédit.

2.1. Analyse des valeurs manquantes

L'inspection systématique du jeu de données montre que la quasi-totalité des variables ne présente aucune valeur manquante. Une seule variable est concernée :

- **taux_endettement** : 983 valeurs manquantes, soit **0,96 %** de la base initiale.

Afin de garantir la cohérence des analyses, l'ensemble de ces observations incomplètes a été supprimé. Cette étape de nettoyage a permis de travailler sur une base entièrement renseignée pour la variable la plus critique du profil client.

2.2. Détection des valeurs aberrantes

Au-delà des valeurs manquantes, plusieurs incohérences ont été détectées dans les données. Les plus significatives concernent la variable **taux_endettement**. En effet, avant nettoyage, certaines observations présentaient :

- des taux **négatifs**,
- des taux **supérieurs à 100 %**, parfois très élevés (jusqu'à plus de 900 %).

Ces valeurs sont économiquement impossibles et traduisent nécessairement des erreurs de saisie ou d'importation. Un nettoyage complémentaire a donc été appliqué, consistant à retirer toute observation avec un taux d'endettement strictement inférieur à 0 % ou supérieur à 100 %**.**

Cette opération permet de recentrer la distribution du taux d'endettement sur une plage réaliste, habituellement comprise entre 10 % et 45 % pour des emprunteurs solvables, tout en conservant la majorité des observations valides.

2.3. Incohérences structurelles et anomalies supplémentaires

D'autres anomalies structurelles ont également été identifiées :

- La variable **mois_demande** présente des valeurs supérieures à 12 (jusqu'à 20), ce qui n'est pas compatible avec un calendrier classique.
- La variable **jour_naissance** peut prendre la valeur 0, ce qui ne correspond pas à une date valide.
- Certains codes postaux sont correctement renseignés mais couvrent une plage très large, incluant potentiellement des territoires ultramarins ou des zones mal saisies.
- Emprunteurs mineurs : Le calcul de l'âge à partir des variables **annee_demande** et **annee_naissance** a révélé la présence de demandeurs ayant moins de 18 ans au moment de la demande, ce qui est légalement impossible pour la souscription d'un crédit à la consommation en France.

Ces incohérences ne sont pas corrigées dans cette première phase, car elles nécessitent des traitements plus complexes (reconstruction des dates, validation géographique). Elles ont toutefois été signalées en tant que **limites importantes** de la base.

2.4. Base finale après nettoyage

Après suppression :

1. des **valeurs manquantes** de taux d'endettement,
2. des **valeurs aberrantes** (<0 % ou >100 %),

La base finale utilisée pour l'analyse descriptive contient :

- **100 384 observations**,
- **18 variables**,
- **0 % de valeurs manquantes** sur l'ensemble des variables quantitatives utilisées.

Nous avons choisi de supprimer ces observations, car elles représentent moins de 2 % de l'ensemble des données. Ce jeu de données nettoyé constitue un socle fiable pour la suite des analyses.

3. Analyse descriptive

L'objectif est d'examiner les variables qualitatives et quantitatives afin d'en dégager les tendances générales, d'évaluer la structure du portefeuille et de préparer les analyses plus avancées à venir.

3.1. Analyse des variables qualitatives

3.1.1. Objet du véhicule et type du véhicule

L'immense majorité des véhicules financés sont des **véhicules particuliers (VP)**, représentant environ 98,8 % du portefeuille, contre seulement 1,2 % pour les véhicules utilitaires (**VU**). Cette répartition témoigne d'un marché principalement orienté vers le financement des besoins individuels plutôt que professionnels.

Concernant le **type de véhicule**, on observe une répartition relativement équilibrée entre :

- **Véhicules neufs (VN)** : 57 367 dossiers (≈ 57 %)
- **Véhicules d'occasion (VO)** : 44 093 dossiers (≈ 43 %)

3.1.2. Type de prêt

Trois types de financements sont représentés :

- **LOA (Location avec Option d'Achat)** : ≈ 56 %
- **VAC (Crédit classique)** : ≈ 36 %
- **LLD (Location Longue Durée)** : ≈ 8 %

La **LOA** est majoritaire, ce qui confirme l'importance croissante des financements locatifs dans les stratégies commerciales des constructeurs et des organismes financiers.

3.1.3. Classe de score

La variable **classe_de_score** est fortement déséquilibrée : la classe **01** concentre à elle seule plus de 80 % des observations. Les classes plus risquées (06 à 09) ne représentent qu'une minorité du portefeuille.

Ce déséquilibre suggère que l'entreprise finance essentiellement des clients considérés comme **faiblement risqués** selon son modèle interne de scoring.

3.1.4. Décision de la banque et décision du client

La banque octroie la majorité des demandes :

- **Octroyé : 89 %**
- **Refusé : 11 %**

Du côté des clients, on observe :

- **Accord client : 85 %**
- **Refus client : 12 %**
- **Sans suite : 3 %**

Ces résultats confirment un taux d'acceptation élevé, à la fois du côté de la banque et du côté des clients, traduisant un portefeuille globalement favorable.

3.2. Analyse des variables quantitatives

3.2.1. Prix d'achat et montant du prêt

Les caractéristiques financières montrent une distribution cohérente avec le marché du financement automobile.

- **Prix d'achat (médiane : 24 520 €)**
  Les véhicules financés appartiennent majoritairement au segment milieu-de-gamme, avec une concentration des prix entre 17 000 € et 33 000 €.
- **Montant du prêt (médiane : 18 887 €)**
  Le montant du financement ne couvre pas systématiquement le prix total du véhicule, ce qui indique une présence non négligeable d'apports ou d'anciennes reprises.

3.2.2. Durée du prêt

La durée du financement se situe principalement entre **48 et 60 mois**, avec une médiane à **49 mois**.
Cela correspond aux standards du marché pour les véhicules neufs et d'occasion récents.

3.2.3. Taux d'intérêt

La distribution du taux d'intérêt présente une forte concentration autour de **0 %**, traduisant des campagnes commerciales attractives.

Cependant, l'amplitude élevée (jusqu'à 11,5 %) montre que certains dossiers sont soumis à des conditions tarifaires plus traditionnelles.

3.2.4. Taux d'endettement

Après nettoyage, la distribution du taux d'endettement est parfaitement cohérente :

- Min : **0,299 %**
- Médiane : **28,405 %**
- Moyenne : **30,857 %**
- Max : **99,991 %**

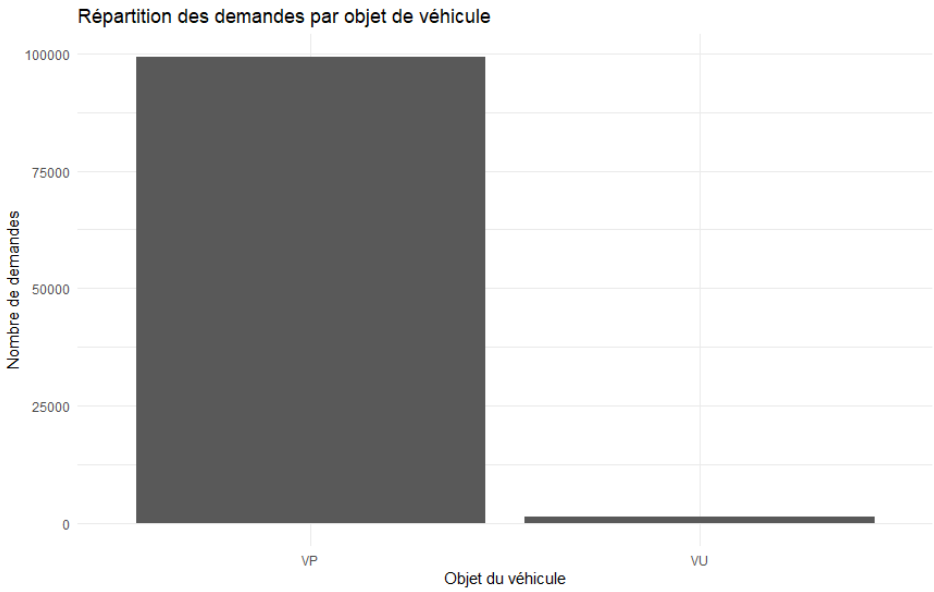
La majorité des emprunteurs présente un niveau d'endettement compris entre **16 % et 42 %**, correspondant aux standards habituels pour des financements automobiles. La suppression des valeurs aberrantes a permis d'obtenir une distribution réaliste pour l'analyse du risque.

3.3. Interprétation des principaux graphiques

Les visualisations suivantes permettent de dégager les tendances majeures du portefeuille de demandes de financement automobile.

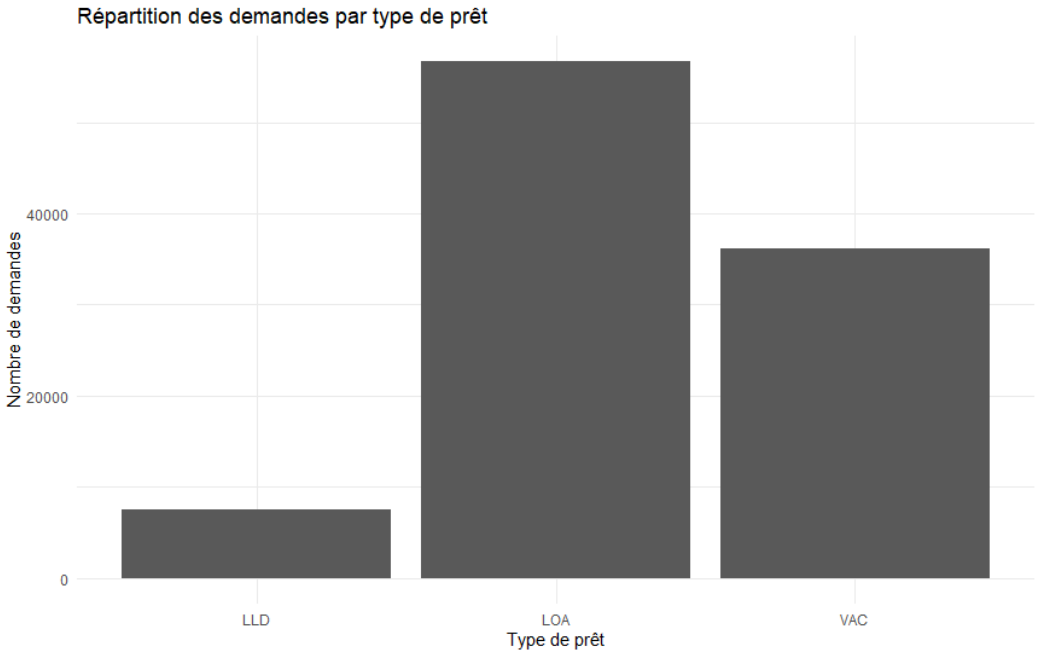
1. Objet du véhicule

La répartition des demandes montre une **quasi-exclusivité des véhicules particuliers (VP)**, les véhicules utilitaires représentant une part marginale.



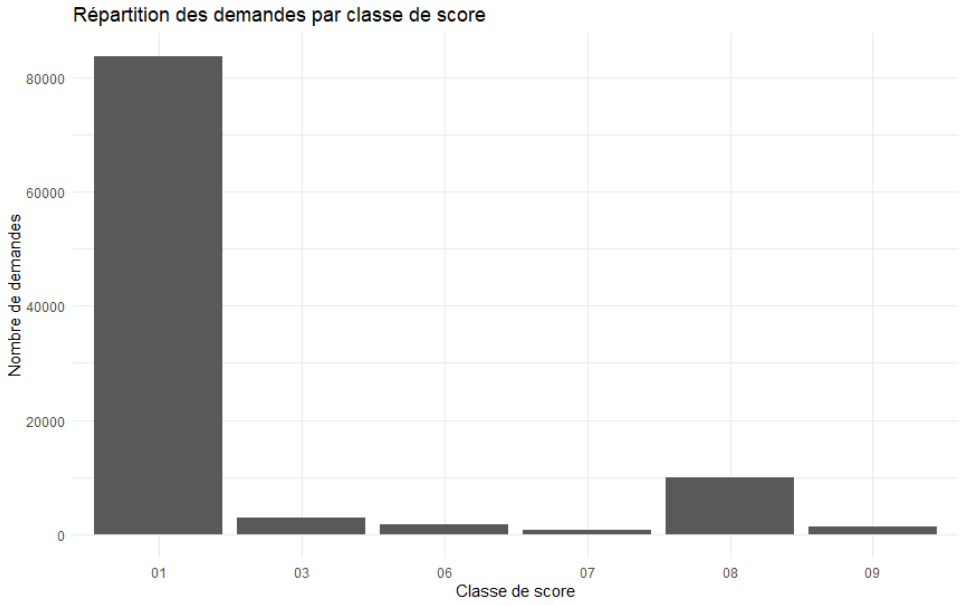
2. Type de prêt

La **LOA** apparaît comme le type de financement dominant, suivie par le crédit classique (VAC). La LLD reste minoritaire.



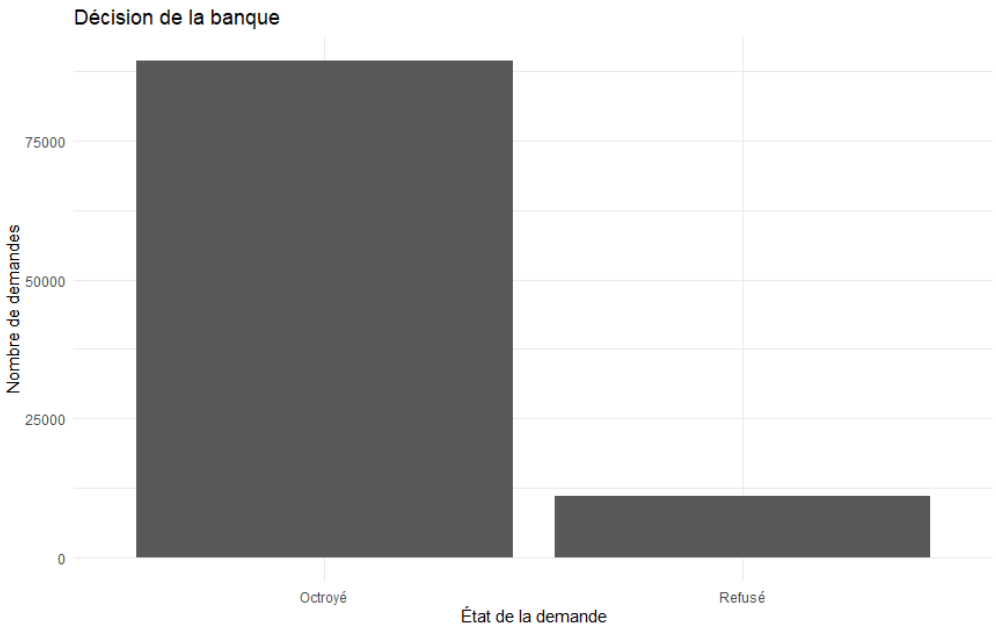
3. Classe de score

La classe de score **01** concentre l'essentiel des demandes, illustrant un portefeuille composé majoritairement de profils considérés comme **faiblement risqués**.



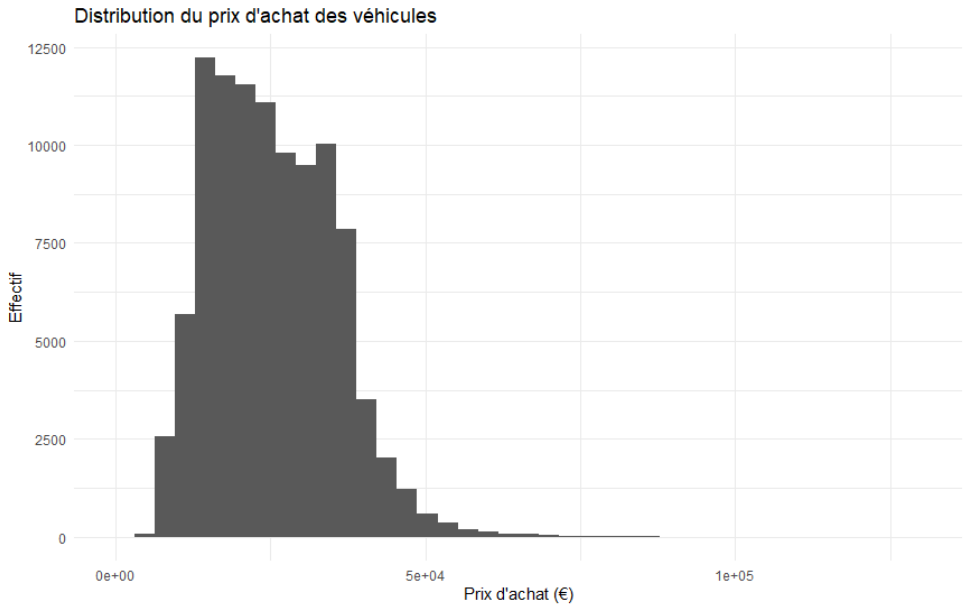
4. Décision de la banque

La banque accepte environ **9 demandes sur 10**, le taux de refus étant relativement faible.



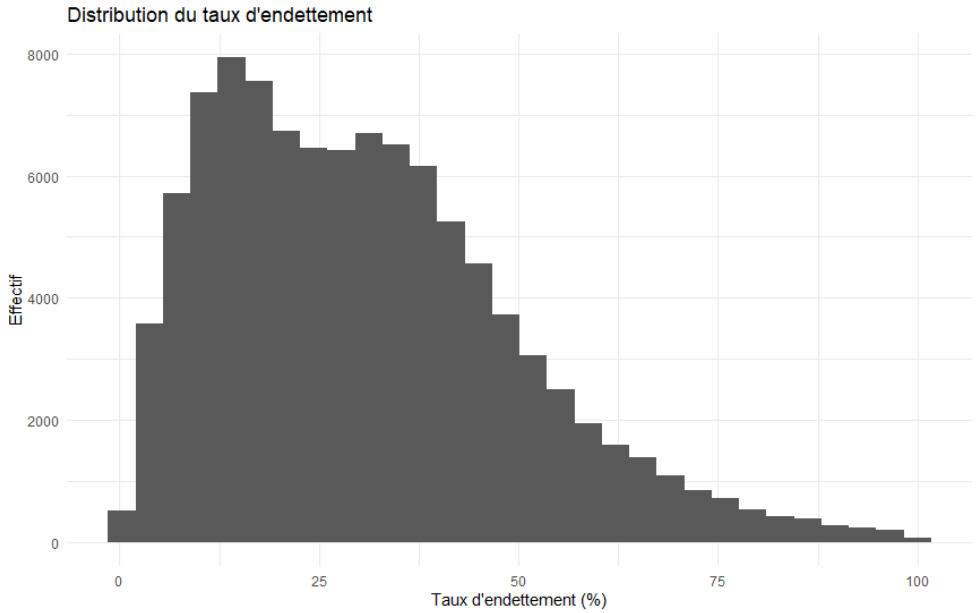
5. Distribution du prix d'achat

Les prix des véhicules se concentrent entre **15 000 € et 35 000 €**, avec une décroissance progressive pour les véhicules plus chers.
→ Le portefeuille est majoritairement constitué de véhicules milieu de gamme.



6. Distribution du taux d’endettement

Après nettoyage, le taux d’endettement se concentre entre **15 % et 45 %**, valeurs cohérentes avec les standards du marché.



4. Présentation de l'outil d'aide à la décision

Au-delà de l'analyse descriptive statique présentée précédemment, le tableau de bord interactif développé sous Streamlit a été conçu comme un outil opérationnel d'aide à la décision. Il ne se contente pas de restituer l'historique des données ; il vise à répondre à deux impératifs stratégiques majeurs pour l'établissement financier : le pilotage de l'activité commerciale et la surveillance de la politique de risque.

4.1 Pilotage de l'activité commerciale et de la production

Le premier axe du tableau de bord est de permettre un suivi dynamique de la performance commerciale. À travers les indicateurs de volume (nombre de demandes, montant total financé) et l'analyse temporelle, l'outil permet de :

- **Suivre la dynamique de production :** Identifier les tendances saisonnières et l'évolution mensuelle des demandes pour anticiper les charges de travail ou les besoins de refinancement.
- **Analyser le mix-produit :** Comprendre la répartition du portefeuille entre les différents produits (prédominance de la LOA, parts de marché du VN vs VO) afin d'ajuster les campagnes marketing.
- **Mesurer la transformation commerciale :** Le suivi des décisions clients ("Accord client" vs "Refus client" ou "Sans suite") permet d'évaluer l'attractivité des offres proposées.

4.2 Pilotage de la politique de risque et de l'octroi

Le second axe, tout aussi critique, concerne la maîtrise du risque de crédit. Le tableau de bord offre une vision détaillée de la qualité des emprunteurs entrants, permettant de :

- **Surveiller la sélectivité de la banque :** Le suivi du **taux d'octroi** (globalement à 89 % sur la période étudiée) permet de vérifier si la politique d'acceptation est en phase avec l'appétence au risque de l'établissement.
- **Contrôler la solvabilité du portefeuille :** Les indicateurs sur le **taux d'endettement moyen** (30,9 %) et la répartition par **classe de score** assurent que la production nouvelle reste dans des standards de risque acceptables (concentration sur les profils à faible risque).
- **Détecter les dérives :** La visualisation des taux d'endettement par classe de score (via les boîtes à moustaches) permet d'identifier rapidement d'éventuelles incohérences ou des segments de clientèle plus fragiles.

5. Conclusion

L'objectif de ce projet était d'analyser un portefeuille de financements automobiles afin de mieux comprendre le profil des clients financés, la structure des prêts et les principaux déterminants du risque de crédit. À partir d'une base initiale de 102 443 demandes, un travail rigoureux de préparation des données a été mené, incluant l'identification des valeurs manquantes, la détection des valeurs aberrantes et la mise en cohérence des principales variables. Après nettoyage, la base finale utilisée pour l'analyse descriptive comprend 100 384 observations et 18 variables, sans valeur manquante sur les indicateurs clés.

Les résultats descriptifs mettent en évidence plusieurs caractéristiques fortes du portefeuille. D'abord, l'activité est très largement orientée vers les véhicules particuliers, les véhicules utilitaires ne représentant qu'une part marginale des dossiers. Le type de financement dominant est la LOA, ce qui confirme le développement des solutions locatives sur le marché automobile. La distribution des classes de score montre que la grande majorité des clients appartient à la classe 01, considérée comme peu risquée. Cette structure explique en partie le taux d'octroi très élevé : la banque accepte environ neuf demandes sur dix, et la plupart des clients donnent leur accord à la proposition de financement.

Les indicateurs financiers sont globalement cohérents avec les pratiques du marché. Les prix d'achat se concentrent sur des véhicules de milieu de gamme, tandis que les montants financés suggèrent la présence d'apports ou de reprises. Le taux d'endettement, après suppression des valeurs incohérentes, se situe majoritairement entre 15 % et 45 %, niveau compatible avec des standards de solvabilité raisonnables. La distribution des taux d'intérêt révèle enfin une coexistence de campagnes promotionnelles à 0 % et de tarifications plus classiques, reflétant une stratégie commerciale différenciée selon les profils ou les produits.

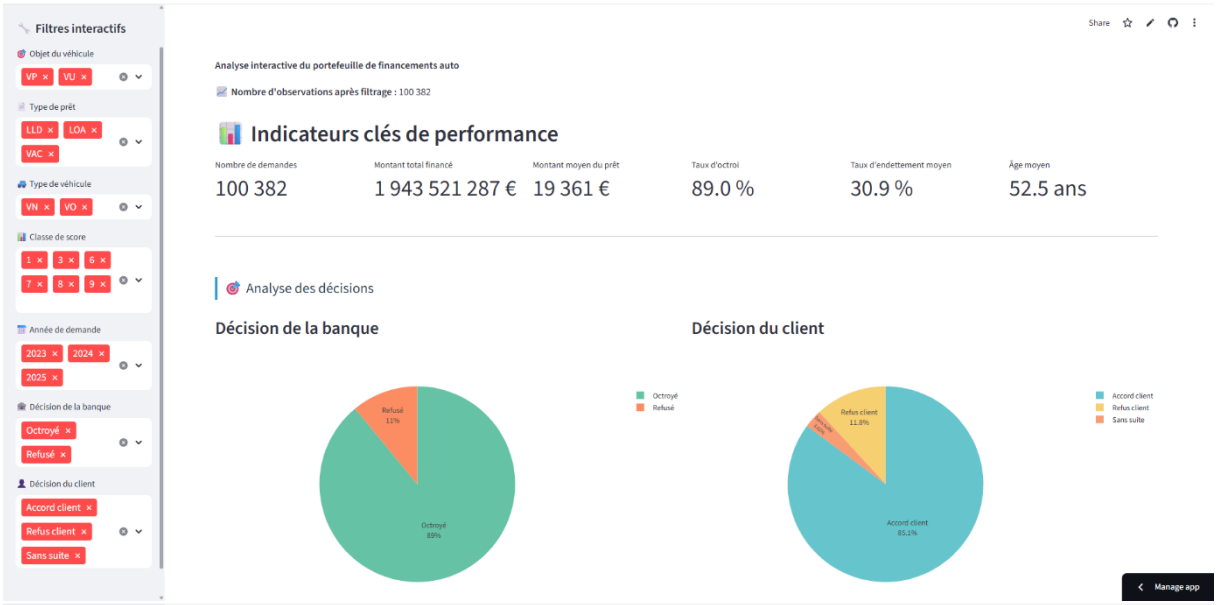
Ce travail présente néanmoins certaines limites. D'une part, l'analyse est essentiellement descriptive et ne permet pas, à ce stade, de mesurer finement l'impact de chaque variable sur le risque de défaut. D'autre part, quelques incohérences résiduelles subsistent (dates approximatives, codes postaux dispersés) et pourraient être corrigées dans le cadre d'un enrichissement ultérieur de la base. Malgré ces limites, l'étude fournit une vision d'ensemble

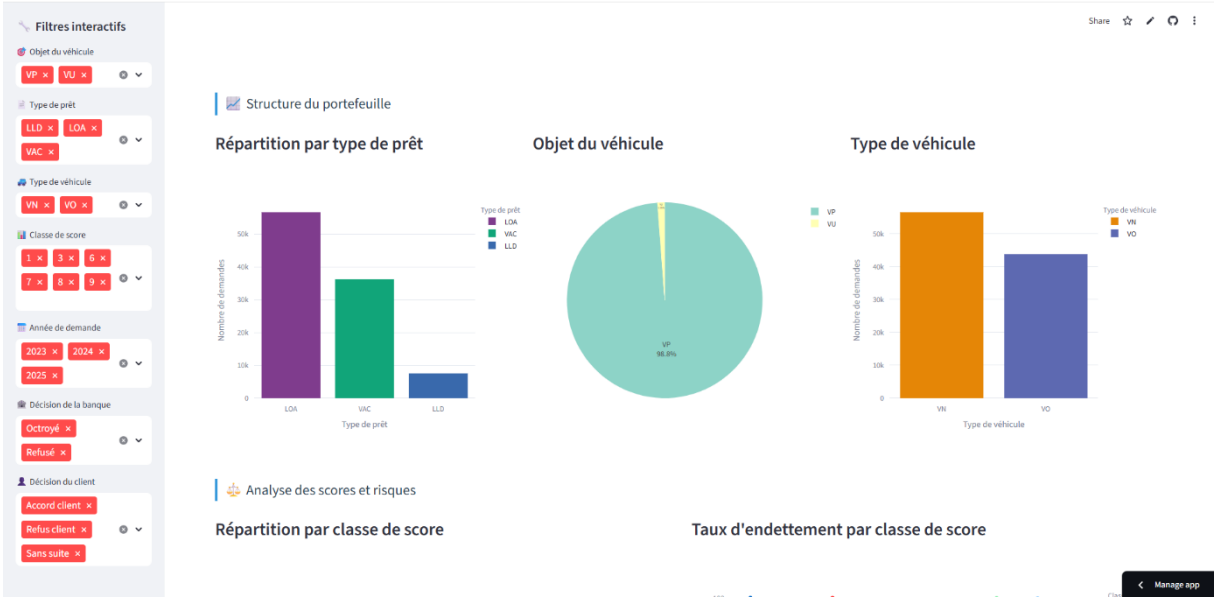
claire et structurée du portefeuille, et constitue une base solide pour des analyses prédictives ou des études de rentabilité plus poussées.

Dans la continuité de ce rapport, la mise en place d'un tableau de bord interactif sous Streamlit permettra de valoriser ces résultats auprès des décideurs. Un tel outil offrira la possibilité de filtrer dynamiquement le portefeuille (par type de prêt, type de véhicule, année ou classe de score), de suivre des indicateurs clés (taux d'acceptation, répartition des montants financés, niveaux d'endettement) et d'explorer les données de manière plus opérationnelle. Le travail de nettoyage et d'analyse réalisé ici constitue donc une étape préparatoire indispensable à la construction de ce futur outil d'aide à la décision.

Tableau de bord (Streamlit)

Lien pour accéder au tableau de bord : https://creditsauto-t9vchhut6e6yxnhh4b39px.streamlit.app/





Filtres interactifs

Objet du véhicule
VP × VU ×

Type de prêt
LLD × LOA × VAC ×

Type de véhicule
VN × VO ×

Classe de score
1 × 3 × 6 × 7 × 8 × 9 ×

Année de demande
2023 × 2024 × 2025 ×

Décision de la banque
Octroyé × Refusé ×

Décision du client
Accord client × Refus client × Sans suite ×

Share

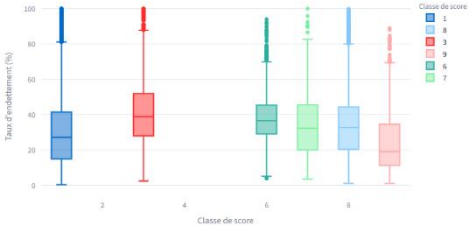
Type de prêt

Type de véhicule

Analyse des scores et risques

Répartition par classe de score

Taux d'endettement par classe de score



Analyse financière

Distribution du prix d'achat

Distribution des taux d'intérêt

Manage app

Filtres interactifs

Objet du véhicule
VP × VU ×

Type de prêt
LLD × LOA × VAC ×

Type de véhicule
VN × VO ×

Classe de score
1 × 3 × 6 × 7 × 8 × 9 ×

Année de demande
2023 × 2024 × 2025 ×

Décision de la banque
Octroyé × Refusé ×

Décision du client
Accord client × Refus client × Sans suite ×

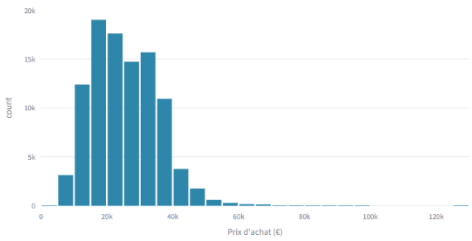
Share

Classe de score

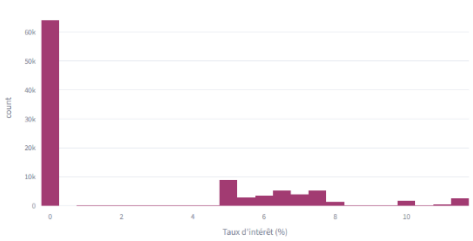
Classe de score

Analyse financière

Distribution du prix d'achat



Distribution des taux d'intérêt



Analyse temporelle

Évolution mensuelle des demandes

Distribution des durées de prêt

Manage app

Filtres interactifs

Objet du véhicule
VP × VU ×

Type de prêt
LLD × LOA × VAC ×

Type de véhicule
VN × VO ×

Classe de score
1 × 3 × 6 × 7 × 8 × 9 ×

Année de demande
2023 × 2024 × 2025 ×

Décision de la banque
Octroyé × Refusé ×

Décision du client
Accord client × Refus client × Sans suite ×

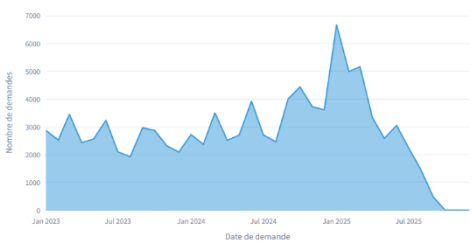
Share

Prix d'achat (€)

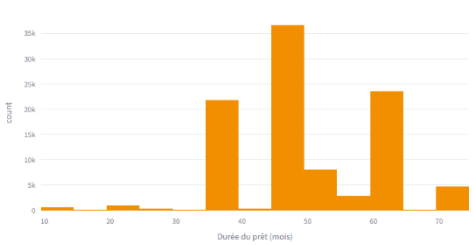
Taux d'intérêt (%)

Analyse temporelle

Évolution mensuelle des demandes



Distribution des durées de prêt



Analyse géographique

Top 15 des départements par nombre de demandes

Manage app

Filtres interactifs

Objet du véhicule
VP × VU ×

Type de prêt
LLD × LOA × VAC ×

Type de véhicule
VN × VO ×

Classe de score
1 × 3 × 6 × 7 × 8 × 9 ×

Année de demande
2023 × 2024 × 2025 ×

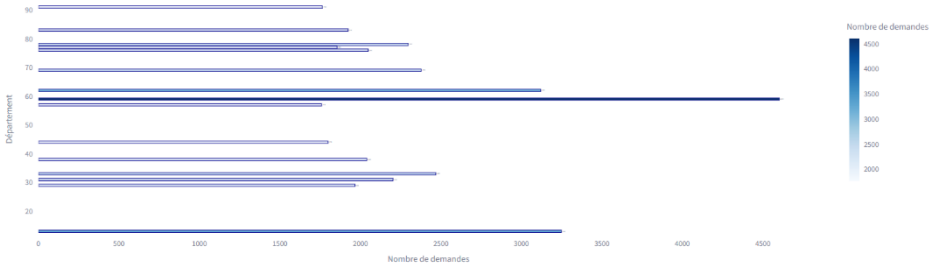
Décision de la banque
Octroyé × Refusé ×

Décision du client
Accord client × Refus client × Sans suite ×

Share

Analyse géographique

Top 15 des départements par nombre de demandes



Données détaillées

Échantillon des données filtrées

	objet_vehicule	classe_de_score	taux_interet	prix_achat	montant_pret	duree_pret	type_vehicule	type_pret	taux_endettement	code_postal	annee_demande	mois_demande	jour_demande	annee_naissance	mois_n
0	VP	1	4.906	28104	8000	36	VO	VAC	19.669	16350	2023	1	28	1957	

Manage app

Filtres interactifs

Objet du véhicule
VP × VU ×

Type de prêt
LLD × LOA × VAC ×

Type de véhicule
VN × VO ×

Classe de score
1 × 3 × 6 × 7 × 8 × 9 ×

Année de demande
2023 × 2024 × 2025 ×

Décision de la banque
Octroyé × Refusé ×

Décision du client
Accord client × Refus client × Sans suite ×

Share

Données détaillées

Échantillon des données filtrées

	objet_vehicule	classe_de_score	taux_interet	prix_ac	montant_pret	duree_pret	type_vehicule	type_pret	taux_endettement	code_postal	annee_demande	mois_demande	jour_demande	annee_naissance	mois_n
0	VP	1	4.906	28104	8000	36	VO	VAC	19.669	16350	2023	1	28	1957	
1	VP	1	4.906	12590	9500	60	VO	VAC	54.979	13009	2023	1	4	1975	
2	VP	1	3.288	19190	8000	60	VO	VAC	13.2	31700	2023	1	28	1948	
3	VP	1	4.908	17354	7354	60	VO	VAC	17.176	29290	2023	2	15	1953	
4	VU	8	4.907	21472.49	15000	60	VN	VAC	12.364	95570	2023	1	6	1997	
5	VP	1	4.907	11576	7000	60	VO	VAC	7.06	13360	2023	1	3	1945	
6	VP	1	4.907	16481	14000	60	VO	VAC	42.27	12700	2023	1	6	1977	
7	VP	8	4.918	13936.6	14288.6	72	VN	VAC	6.584	69560	2023	1	19	1943	
8	VP	1	4.907	12490	12778	58	VO	VAC	38.399	56410	2023	1	10	1966	
9	VP	3	4.906	8784.24	8990	72	VO	VAC	39.502	85000	2023	1	10	2000	

Affichage de 500 lignes sur 100,382 au total. Utilisez le bouton de téléchargement pour obtenir toutes les données.

Statistiques descriptives des données filtrées

Télécharger les données filtrées (CSV)

Tableau de bord créé avec Streamlit • Données issues de l'analyse descriptive de 100 384 demandes de crédit auto

Manage app