

Université Paris-Nanterre

SANTANDER CONSUMER BANQUE

DEPARTEMENT DES RISQUES

Ingénierie Statistique et Economique de la Finance, de l'assurance et du Risque

Modélisation Supervisée et Gestion du Risque : Optimisation des stratégies de Décision

RAPPORT D'ALTERNANCE

Rédigé par : Ayman YAYA

Tuteur : Rafael MILANI

Responsable d'alternance : Patrice BERTAIL

Le 05 septembre 2024



Table des matières

REMERCIEMENTS

1 Introduction générale

- 1.1 Introduction
- 1.2 Introduction aux Risques
- 1.3 Rôle des mathématiques appliquées et de l'informatique dans le Risque
- 1.4 Objectifs
- 1.5 Rappel du Contexte et le Plan

2 Présentation de l'entreprise

- 2.1 Le groupe Santander
 - 2.1.1 Histoire
 - 2.1.2 Activités et Valeurs
- 2.2 Santander Consumer Banque
 - 2.2.1 Histoire SCB
 - 2.2.2 Activités
 - 2.2.3 Objectifs
- 2.3 Direction des risques
 - 2.3.1 Structure et Fonctionnement
- 2.4 Département Supervision des Risques
 - 2.4.1 Types de risques
 - 2.4.2 Structure

3 La gestion du risque de crédit

- 3.1 Types de risques de crédit
- 3.2 Définition du défaut
- 3.3 Le processus de gestion du crédit de risque
- 3.4 Situation Actuelle

4 Description des missions

- 4.1 Déroulement
- 4.2 Outils
- 4.3 Construction et Automatisation des Reportings
 - 4.3.1 Data Quality Checks
 - 4.3.2 MIS

5 Analyses relatives aux SDS changes

- 5.1 Définitions préliminaires
- 5.2 Rappel de la stratégie d'octroi
- 5.3 Changement de la Cut-off du Scoring Particulier
 - 5.3.1 Contexte
 - 5.3.2 Analyse de scénarios
 - TESLA (B2C)
 - 5.3.3 Tests de régression et de non-régression
 - 5.3.4 Calcul et analyse de l'impact

6 Projets

- 6.1 Les besoins et mes propositions
 - 6.2.1 Projet Fraude
 - 6.2.2 Projet PBI
- 6.2 Production Projet Fraude
 - 6.2.3 Collecte des données

6.2.4	Traitements des données	32
6.2.5	Impact du projet de Fraude	33
6.3	Production Projet PBI	33
6.3.1	Collecte des données	33
		35
6.3.2	Traitements des données	36
6.3.3	Qualité des données	37
6.3.4	Modèle et Mesures	37
7	Conclusion	39
8	Bibliographie	40
9	Annexes	41

REMERCIEMENTS

En préambule à ce rapport, je souhaite adresser mes remerciements aux personnes ayant contribué au bon déroulement de cette alternance de fin d'études.

Je tiens tout d'abord à exprimer ma sincère gratitude envers mon tuteur d'alternance, Rafael Milani, Analyste Risque de crédit chez Santander Consumer Banque, pour sa grande disponibilité à mon égard, sa bienveillance et son soutien constant. Ses enseignements ont été essentiels à mon apprentissage du métier et à ma progression tout au long de la mission.

Mes remerciements vont également à Joachim Priou, Manager de l'équipe de Modélisation et de Statistiques et de Retail Risk pour m'avoir accueilli chaleureusement au sein de l'équipe et m'avoir impliqué dans les projets en cours. Ses encouragements furent souvent une source d'inspiration et de motivation.

Je remercie l'ensemble de l'équipe des risques avec qui j'ai collaboré pour m'avoir offert un environnement de travail aussi harmonieux et agréable.

Enfin, je remercie l'Université Paris-Nanterre et en particulier les responsables de la filière ISEFAR, qui ont rendu cette expérience possible.

1 Introduction générale

1.1 Introduction

Dans le cadre de mon alternance de fin d'études à l'Université Paris-Nanterre en ISEFAR, j'ai eu l'opportunité d'intégrer le département des risques de **Santander Consumer Banque** à Levallois-Perret en Île-de-France en tant qu'**Analyste Risque Data sur la partie Crédit**.

Santander Consumer Banque est la succursale française de **Santander Consumer Finance**, une division du groupe bancaire espagnol Santander, spécialisée dans les services financiers aux consommateurs et aux entreprises. Avec une présence dans 15 pays européens, Santander Consumer Finance est un leader sur son secteur en Europe, comptant plus de 15 500 collaborateurs, 21 millions de clientset plus de 130 000 points de vente.

Depuis le début de son activité commerciale en France en **janvier 2016**, Santander Consumer Banque France s'est positionnée sur le marché du **financement automobile français**, troisième marché automobile européen, en offrant des produits de financement et des services associés. Cependant, l'entreprise conserve une forte ambition de développement vers d'autres marchés dans le but de s'imposer parmi les leaders du crédit à la consommation en France.

1.2 Introduction aux Risques

Le **Risque** est défini comme un « *Danger éventuel prévisible* ». C'est la probabilité que des individus, des biens ou des organisations soient affectés par un danger plus ou moins reconnu. Il est donc caractérisé par une **incertitude**.

Le risque financier, comme le dit Pierre Conso, « *est omniprésent et multiforme [...]* *Le combattre concerne donc tous les acteurs* ». Face à celui-ci, les institutions financières doivent se mobiliser massivement dans l'intérêt général de leur santé.

Le **risque de crédit**, défini dans le cadre de Bâle III comme « *le risque de perte économique résultant du non-respect par la contrepartie de ses obligations contractuelles* », est la forme la plus courante de risque de contrepartie.

Dans l'activité de prêts, le risque de crédit peut prendre plusieurs formes, notamment le **risque de défaut**. C'est la probabilité qu'un emprunteur ne parvienne pas à rembourser sa dette conformément aux conditions convenues. Cette situation d'**insolvabilité** engendre des pertes partielles ou totales des créances que les provisions doivent couvrir, et résulte en une diminution des bénéfices.

En vue de l'environnement très concurrentiel et des taux d'intérêts extrêmement bas, il devient de plus en plus difficile de s'imposer sur le marché du crédit. De plus, les banques demeurent vulnérables aux chocs économiques majeurs. Par exemple, la crise de 2008 a souligné les faiblesses des institutions bancaires face à ce risque. Elle a surtout mis en évidence la nécessité de renforcer les pratiques de gestion du risque de crédit pour éviter de telles conséquences à l'avenir.

Désormais, une gestion stratégique du risque est considérée comme une force majeure. L'application correcte d'une bonne politique permet d'augmenter considérablement ses rentabilités, de se différencier des concurrents tout en préservant sa santé financière.

1.3 Rôle des mathématiques appliquées et de l'informatique dans le Risque

La gestion du risque de crédit chez Santander Consumer Banque comprend : **l'évaluation** des emprunteurs, **l'analyse** du risque, la prise de décision de prêt (**l'octroi**), **la surveillance** continue (Risk Oversight) et **l'atténuation** des pertes en cas de défaut. Les mathématiques et l'informatique interviennent à chaque étape de ce processus.

Pour revenir sur la notion d'incertitude, prédire la probabilité qu'un événement se produise peut s'avérer difficile. Toutefois, si suffisamment de données historiques relatives à cet événement existent, c'est possible. C'est ainsi que des modèles statistiques sont construits pour prédire **la probabilité de défaut** des emprunteurs. Ces modèles fournissent aussi le **Scoring**, outil d'aide à la prise de décision pour les analystes.

En cas d'impayés, la banque n'est jamais totalement sûre de récupérer son investissement. L'analyse de données contribue à la prévention en **identifiant des tendances émergentes, des corrélations entre certaines caractéristiques de l'emprunteur et sa solvabilité**. Cela permet de mieux **calibrer le modèle de Scoring** et de minimiser les pertes.

Ensuite, la surveillance du portefeuille se fait par le biais **d'indicateurs calculés à partir des données relatives au risque (retards d'impayés, échéances impayées, dégradations dans l'admission, etc)**.

C'est l'essentiel du métier de Risk Data Analyst chez Santander Consumer Banque.

1.4 Objectifs

Cette alternance de fin d'études permet de mettre en pratique mes connaissances théoriques acquises tout au long du cursus académique, notamment les connaissances en statistiques et en modélisation financière. C'est aussi l'occasion de se familiariser avec l'environnement professionnel et les attentes du monde du travail, tout en développant des compétences transversales telles que la communication, le travail en équipe, la gestion de projet et le respect des dates limites. Le but ultime étant de renforcer l'expertise dans le domaine financier et d'affiner ses intérêts pour s'orienter vers un choix de carrière déterminant

L'objectif de mon alternance en tant que Data Analyste Risque se résume en **la participation à la surveillance et à l'analyse du comportement du risque de crédit, ainsi qu'à l'évolution du système décisionnel risque de Santander Consumer Banque. Ce rapport se concentre sur les efforts déployés pour améliorer l'efficacité de ce système décisionnel, notamment par l'intégration d'analyses et d'outils de surveillance automatisés, afin de renforcer la capacité de la banque à gérer les risques de manières proactive et efficace.**

1.5 Rappel du Contexte et le Plan

Santander Consumer Banque est la succursale française de Santander Consumer Finance, une division du groupe bancaire espagnol Santander. Fondée en **2014**, SCB est spécialisée dans **les produits de financement liés à l'automobile, offrant des solutions de crédit, de crédit-bail (leasing) et d'assurances destinées à une clientèle de particuliers et d'entreprises** partout en France métropolitaine. Les produits sont commercialisés en **B2B par l'intermédiaire des concessionnaires**, et en **B2C** ou en "vente à distance" pour la clientèle **TESLA**. SCB propose aussi des prêts personnels en ligne en partenariat avec **La Centrale**.

Ce rapport commence par une présentation de SCB, retraçant son historique et ses activités principales, avec un focus particulier sur le fonctionnement du département des risques. Cette introduction vise à fournir un contexte essentiel sur la mission de SCB dans la gestion des risques financiers.

La gestion du risque de crédit est ensuite explorée, en décrivant les processus et méthodologies que SCB utilise pour identifier et atténuer les risques liés au crédit, en accord avec les réglementations en vigueur. Les missions effectuées durant l'alternance sont détaillées dans la partie suivante, avec un accent sur l'automatisation des reportings et l'analyse des indicateurs de risque, illustrant l'importance de ces tâches pour l'optimisation des processus internes.

Le rapport se penche ensuite sur les analyses des changements dans les Stratégies de Décision et de Scoring (SDS), évaluant l'efficacité des ajustements et des nouvelles approches mises en place pour améliorer la gestion des risques.

Les projets spécifiques, y compris l'utilisation de Power BI et d'autres initiatives innovantes, sont présentés, montrant leur impact sur l'amélioration des pratiques de gestion du risque de SCB.

Enfin, la conclusion synthétise les résultats de l'alternance.

Abréviations

- **SCF** : Santander Consumer Finance
- **SCB** : Santander Consumer Banque
- **VAC** : Vente à crédit
- **VO** : Véhicule d'occasion
- **VN** : Véhicule Neuf
- **SME** : Small and medium-sized enterprises
- **PART** : Particuliers
- **RO** : Risque Opérationnel
- **BPO** : Bank Payment Obligation
- **TTY** : Time To Yes
- **NPL** : Non-Performant Loans
- **DQ** : Data Quality
- **MIS** : Management Information System

Remarque : Pour des raisons de confidentialité, certains chiffres comme le coût du risque et les bilans des partenariats, ne seront pas communiqués.

2 Présentation de l'entreprise

2.1 Le groupe Santander

Le Groupe Santander est un **leader mondial du financement** fondé en **1875** avec la création de la **Banco Santander** en Espagne. Il s'agit de la 1^{ère} banque espagnole et du 4^e groupe bancaire au monde en termes de capitalisation boursière. Il compte aujourd'hui plus de 202 000 employés, plus de 133 millions de clients et une présence internationale dans plus de 40 pays. Le groupe est coté sur plusieurs bourses, dont celles de Madrid, New York, Londres et Milan.

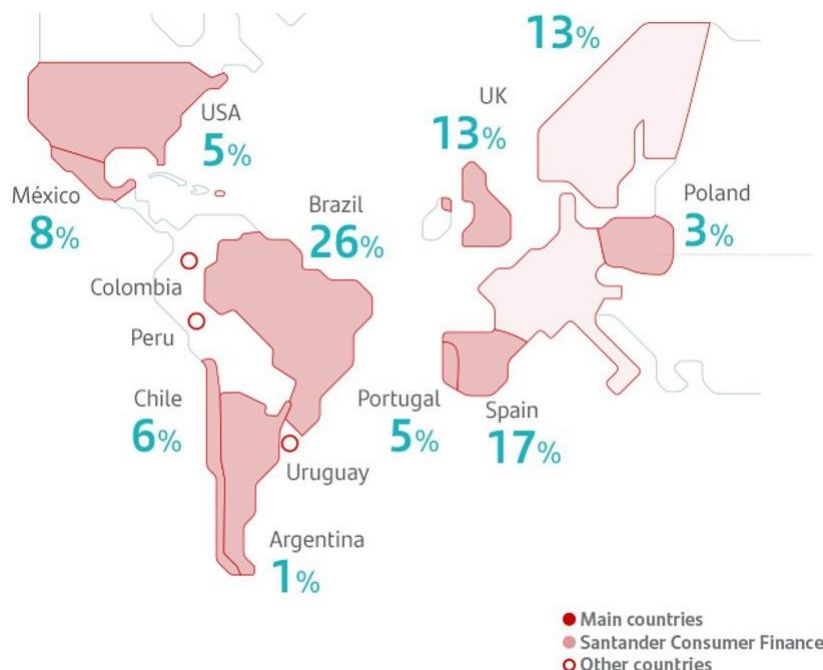


Figure 1 – Présence Internationale du groupe Santander

2.1.1 Histoire

Fondé en 1857 par décret la Reine Isabel II d'Espagne en tant que petite institution locale, la banque a connu une croissance significative, notamment grâce à une série d'acquisitions stratégiques, lui permettant d'étendre sa présence à l'échelle internationale.

En effet, les années 1930-1950 marquent l'expansion à échelle nationale de Banco Santander sous la direction d'Emilio Botín Sanz de Sautuola y López, petit-fils du fondateur de la banque, avec l'acquisition de la Banco de Avila à Madrid en **1942**, et de Banco Mercantil en **1964** principal concurrent de Banco Santander. Au début des années 1960, Santander initie son développement en outre-Atlantique avec les premiers bureaux en Amérique Latine (Cuba, Argentine, Mexique, Venezuela) et en Europe à Londres, qu'il poursuit aux cours des quatre prochaines décennies.

Les années 2000 ont propulsé Santander au rang de leader mondial. En effet, en **1999**, motivé par l'arrivée de l'euro, le Groupe Santander a réalisé une acquisition majeure en fusionnant avec Banco Central Hispano (BCH), une autre grande banque espagnole. Enfin, en **2004**, Santander a fait une autre acquisition significative, Abbey National. Cela lui a permis de s'implanter sur le marché britannique, l'un des marchés financiers les plus importants et de renforcer ses activités de détail.

2.1.2 Activités et Valeurs

Aujourd'hui, le Groupe Santander opère dans quatre domaines financiers principaux : la banque de détail, la banque commerciale, la banque d'investissement et la gestion d'actifs. Il offre une large gamme de services tels que les prêts, les produits d'épargne, les cartes de crédit, les assurances et les services de paiement aux particuliers, aux entreprises et aux institutions. Les clients particuliers représentent tout de même 36% des opérations bancaires. En termes de profit, l'Europe et l'Amérique Latine y contribuent similairement. Le groupe a défini trois valeurs que toutes les filiales partagent :

- **Simple** : Être clairs dans les propositions et faciliter les processus.
- **Personnel** : Privilégier la compréhension, la souplesse et un relationnel attentif.
- **Juste** : Suivre une approche réaliste, honnête ainsi que transparente dans les relations.

2.2 Santander Consumer Banque

Santander Consumer Banque S.A est la succursale de Santander Consumer Finance en France, une division du Groupe Santander spécialisée dans les services financiers aux consommateurs. SCF a été créée en 1997 et est rapidement devenue leader du financement des véhicules sur le continent. Elle compte aujourd'hui 15 000 employés, 18 millions de clients dans 15 pays européens (Allemagne, Autriche, Belgique, Danemark, Espagne, Finlande, Pays-Bas, Italie, Norvège, Pologne, Portugal, Royaume-Uni, Suède, Suisse et France) et rapporte 12.5% de bénéfice au groupe.

2.2.1 Histoire SCB

SCB, fondée en **2014** dans le but de renforcer la présence du groupe, en France. Elle a obtenu son **agrément bancaire de l'ACPR en 2015** et a lancé son activité commerciale en **janvier 2016**. Depuis lors, l'entreprise a connu une croissance constante et regroupe maintenant plus de 120 collaborateurs : une cinquantaine d'employés au sein de la Direction Commerciale et marketing, dont 2/3 sur le terrain, et une trentaine de collaborateurs au sein de la Direction Risques et Recouvrement. La société exerce son activité en France (hors Monaco et Outre-Mer) et l'ensemble de ses opérations est libellé en euros.

En sept ans, SCB est devenu un acteur référent du financement de véhicules en France, notamment grâce à **l'acquisition de 50% de Banque PSA France (anciennement SOFIB) en février 2015**. Les différents partenariats commerciaux développés au cours des dernières années ont catalysé cette croissance (KTM en 2018, Piaggio en 2023, etc). **2021** fut une année clé puisqu'elle marque la mise en place d'un nouveau partenariat entre SCB et le constructeur américain de voitures électriques, **TESLA**, ainsi que l'introduction du canal de vente à distance B2C, où 100% des ventes se font en ligne, sans avoir recours aux concessionnaires. Grâce à tous les nouveaux partenaires développés, le chiffre d'affaires a progressé de 28% par rapport à 2020, alors que les effectifs n'ont augmenté que de quelques employés.

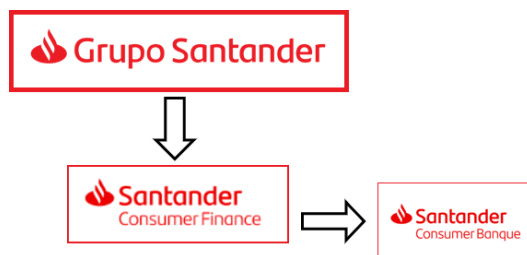


Figure 2 – Organigramme de Santander Consumer Banque (Annexe I)

2.2.2 Activités

SCB concentre son activité autour des produits de **financement automobile**, par crédit, crédit-bail et location avec option d'achat, destinés à une clientèle de particuliers et d'entreprises. Ceci comprend le financement de l'acquisition de **véhicules neufs et d'occasion : Auto, Moto et VDL (Véhicule de loisirs)**. D'autres services sont aussi proposés comme les **assurances non-vie** du capital financier en cas de vol ou de destruction totale du véhicule et les **assurances-vie** portant sur l'emprunteur (décès, perte d'emploi, invalidité, etc). (Description plus détaillée de chaque produit en Annexe II).

	Crédit				Leasing	
	VAC (Vente à crédit)	CBAILC (Crédit Bail)	CBAILB (Crédit Bail Ballon)	Crédit à Stock	LOAC (Location avec Option d'Achat)	LOAB (Location avec Option d'Achat Ballon)
Durée de financement	12 à 156 mois	36 à 60 mois	24 à 61 mois		36 à 60 mois	24 à 61 mois
Véhicules éligibles	Véhicules Neufs/ Véhicules Occasion				VN/VO	
Clients éligibles	Particuliers / Entreprises	Entreprises	Entreprises	Concessionnaires (Wholesale)	Particuliers	Particuliers

Figure 3 – Offres de financement Automobiles de SCB

Avec sa présence commerciale dans cinq régions de la France (Île-de-France, Nord, Ouest, Est et Sud), les offres de SCB s'adressent aux trois profils suivants :

- **Particuliers (Retail)** : Individus ayant besoin d'un financement pour acheter ou louer un véhicule à titre personnel.
- **Entreprises (Retail)** : Dirigeants d'entreprises ayant besoin d'un financement pour acheter ou louer un véhicule au nom de l'entreprise dans le cadre de leurs activités.
- **Concessionnaires (Wholesale)** : Concessionnaires automobiles ayant besoin d'une ligne de crédit afin de financer leurs stocks de véhicules.

Aujourd'hui, de 70% des encours enregistrés proviennent des crédits VAC destinés aux particuliers, et 70% de dépôts de la clientèle Wholesale.

Les produits sont commercialisés à travers deux canaux de vente, le B2B (ventes par l'intermédiaire des concessionnaires) et le B2C (pour les clients Tesla, Cazoo et Carnext). Les concessionnaires disposent d'un outil de vente B2B multi-support avec une messagerie en ligne et un suivi des dossiers en temps réel. Actuellement, il s'agit d'un des outils les plus modernes du marché, ce qui lui donne un avantage concurrentiel en termes d'innovation technologique et de service à la clientèle.

2.2.3 Objectifs

Le marché du financement automobile français est milieu concurrentiel difficile à pénétrer. En effet, les acteurs majeurs du financement de l'automobile établis depuis plus de 50 ans (**RCI Bank and Services, PSA Banque, BMW, Volkswagen Financial Services, etc**) détiennent 75% du marché. Par conséquent, SCB doit non seulement se démarquer des autres banques indépendantes comme **Cetelem** et **Viaxcel**, mais aussi consolider sa présence parmi les banques captives de marque.

Il faut aussi noter que la crise pandémique, en plus la crise de l'approvisionnement en semi-conducteurs dans le marché automobile, ont nui à la croissance de l'entreprise. En effet, le premier confinement en mars 2020, a énormément ralenti l'activité de SCB. Le stress du marché de la liquidité a aussi contribué à ses difficultés de refinancement. En sortie de confinement les nombreuses et soudaines demandes de crédit étaient difficiles à gérer d'un point de vue opérationnel.

Depuis cette période de stress, SCB a réussi à reprendre la situation en main et espère continuer sa croissance à rythme élevé. Pour s'y faire, des projets d'enrichissement du portefeuille clients et produits ont été mis en place. En plus du partenariat avec TESLA cité précédemment, l'entreprise a lancé en **janvier 2023** un parcours de vente C2C "100% digital" avec **La Centrale**, société spécialisée dans le courtage en crédits à la consommation. Désormais, en visitant leur site en ligne, les utilisateurs ont accès aux offres de crédits personnels de SCB.

2.3 Direction des risques

La direction des risques est le 2^e département le plus grand chez Santander Consumer Banque. Comme dans toute institution financière, ce département est structuré de manière à assurer une gestion efficace et polyvalente des risques liés à l'activité bancaire. Il est actuellement dirigé par Abdou SOW, dirigeant effectif de l'entreprise. (Organigramme de la direction des risques en Annexe III).

La direction des risques se charge de l'identification, la classification, la gestion et la prévention, tout en veillant à ce que les politiques et les procédures de gestion des risques soient respectées. Elle joue un rôle essentiel dans la préservation de la stabilité financière de l'entreprise et dans la protection de ses intérêts.

2.3.1 Structure et Fonctionnement

La direction des risques est composée de cinq équipes spécialisées qui travaillent en collaboration pour atteindre les objectifs fixés par le conseil de direction au quartier général de SCF, à Madrid. Ce principe de séparation des tâches permet de distinguer les missions appartenant aux services "générateurs de risques" et des fonctions de contrôle et de surveillance.

- **Département Acceptation** : aussi appelée l'équipe Retail. Elle s'occupe de la réception, l'analyse et la validation des demandes de financements soumises par des particuliers et des professionnels. L'analyse consiste à examiner les antécédents financiers, la situation professionnelle et les garanties (co-demandeur, versement initiaux, etc).
- **Département Financements Réseau** : aussi appelée l'équipe Wholesale. Elle s'occupe aussi de l'**octroi**, i.e l'évaluation de la **solvabilité financière** des concessionnaires avant de leur accorder des **crédits rotatifs**. C'est une ligne de crédit préapprouvée qui permet au concessionnaire de disposer d'un certain montant pour financer son inventaire de véhicules, avec la flexibilité de pouvoir emprunter et rembourser plusieurs fois au cours de la durée du contrat. Ce type de crédit implique souvent un volume assez élevé, et donc un risque important (de faillite, de fraude, etc), puisque les véhicules financés ne peuvent pas être revendiqués.

- **Département Supervision des Risques** : responsable de la mise en œuvre des systèmes de mesure et de surveillance des **risques financiers** et **non financiers**. Il veille à ce que les procédures de prise de risques respectent dans **l'appétence aux risques**.
- **Département Recouvrements** : responsable de la gestion des encours en impayé en phase amiable ou contentieuse. Comme vu dans la section précédente, les crédits constituent une part importante des revenus de SCB. En récupérant les paiements en retard ou les montants impayés, l'entreprise préserve sa rentabilité et maintient un flux de trésorerie positif.
- **Département Contrôle Interne et Risque opérationnel** : Cette équipe contrôle le respect de règles et procédures au sein des différentes directions de l'entreprise, selon une **Charte de contrôle permanent**. Celle-ci est revue et mise à jour au minimum une fois par an ou suite aux modifications significatives. Elle est ensuite validée auprès du Conseil de Surveillance. L'ensemble des contrôles dont le résultat n'est pas satisfaisant font l'objet d'un plan d'action et d'un suivi par le Département Conformité en collaboration avec le département concerné.

Pour résumer le système de contrôle interne de SCB, il faut considérer **trois lignes de défense** : (Selon l'arrêté du 3 novembre 2014) :

- * Première ligne : les unités opérationnelles. Les Managers surveillent constamment l'activité de leurs équipes, conformément aux instructions du Département Risques non financiers.
- * Deuxième ligne : assuré principalement par le Département des Risques non financiers qui définit le partage des rôles, de responsabilités et de processus de contrôle, en collaboration avec le département Conformité. Le département des risques financiers y participe minoritairement, en établissant un plan de contrôle pour le suivi de l'exposition aux risques financiers.
- * Troisième ligne : la Direction de l'Audit Interne assure un contrôle périodique.

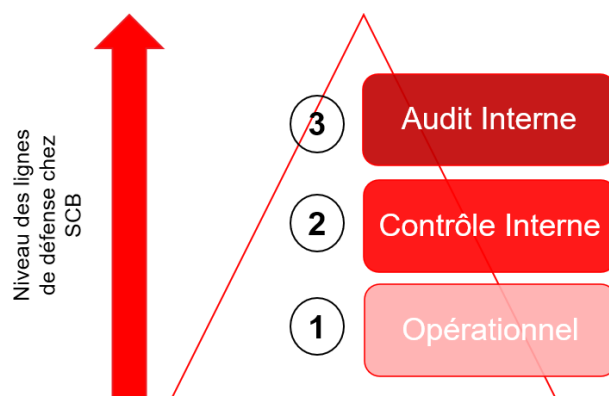


Figure 4 – La défense contre les risques chez SCB

2.4 Département Supervision des Risques

Le Risk Appetite (ou appétence au risque) est le niveau de risque maximum que l'entreprise est prête à accepter pour atteindre ses objectifs. Ce seuil limite est à prendre en compte lors de la prise de décisions concernant les activités commerciales, les politiques de crédit, les modèles de notation, les politiques de provisionnement, etc.

Aujourd'hui, le coût de risque chez SCB continue à baisser grâce à l'augmentation des volumes provenant des partenariats (clientèle premium), mais surtout grâce à la mise en place d'un **nouveau système de gestion de l'octroi en décembre 2020** qui n'a cessé d'évoluer depuis. L'objectif étant d'atteindre le seuil de 1% de pertes liées au risque de crédit en **2026**.

2.4.1 Types de risques

L'appétit au risque doit prendre en compte tous les risques auxquels est exposée SCB, à la fois les risques internes et les risques externes (i.e dommages engendrés des fournisseurs, usines, etc.). Ceux-ci sont soumis à des limites, avec un niveau de tolérance.

	Natures de risques	Criticité brute
Risques financiers	Risque de crédit	
	Risque de marché	
	Risque de liquidité	
	Risques structurels et de capital ¹	
Risques non financiers	Risque opérationnel	
	Risques juridique et de non conformité	
	Risque de conduite	
Risques transverses	Risque de modélisation	
	Risque de réputation	
	Risque de stratégie	

Figure 5 – Types de Risques chez SCB

SCB est confronté principalement à quatre types de risques financiers :

- **Risque de Crédit** : le risque de perte financière due à l'insolvabilité ou à la détérioration de la qualité du crédit d'un client.
- **Risque de marché** : le risque que SCB subisse une perte de valeur économique suite à des modifications des taux d'intérêt dans le Bank-Book, des modifications des taux de change ou des prix des véhicules.
- **Risque de liquidité** : le risque que l'entreprise, tout en étant solvable, ne possède pas suffisamment de liquidités disponibles pour remplir ses obligations au moment requis.
- **Risque structurel et de capital** : le risque de ne pas disposer d'un capital suffisant pour atteindre les objectifs et remplir les obligations réglementaires et les attentes du marché.

Les risques non financiers sont les suivants :

- **Risque opérationnel** : le risque de perte due à des processus, à l'erreur humaine, à des systèmes internes défectueux ou à une cause extérieure (incendies, catastrophes naturelles, etc)
- **Risque de conduite** : le risque résultant d'une inadéquation entre les pratiques ou les produits de la banque et ses clients.
- **Risque réglementaire ou risque de conformité** : le risque de pertes dues au non-respect des codes et de la réglementation.

Quant aux risques transversaux, il peut s'agir de :

- **Risque de modèle** : le risque de pertes dues aux erreurs de modélisation et d'utilisation de ces modèles.
- **Risque de réputation** : le risque de pertes liés à la perception publique de la banque de ses partenaires, ses investisseurs, etc.
- **Risque stratégique** : le risque de mauvaise mise en place des décisions ou l'incapacité de s'adapter aux changements du business.

Comme le montre la figure ci-dessus, les deux risques les plus critiques chez SCB, sont le risque crédit, et le risque de fraude, qui est considéré comme un risque opérationnel, mais aussi comme un risque de crédit.

Une fois les risques identifiés, on les classe en fonction de leur criticité, c'est-à-dire leur gravité et leur probabilité d'occurrence. En fonction de cette mesure, des processus de gestion sont mis en place pour traiter les urgences en priorité, et puis suivre leurs comportements.

Il faut noter que même après avoir appliqué toutes les mesures possibles, on ne peut pas éliminer tous les risques à 100%. On parle alors de **risque résiduel**. Il peut être dû aux erreurs humaines ou à un changement profond dans le contexte économique ou géopolitique ou réglementaire obligeant l'entreprise à restructurer l'intégralité de son système stratégique. Par exemple, suite à la guerre en Ukraine, le groupe Santander a décidé de *ne plus s'engager pas dans de nouvelles affaires avec des entreprises russes*. Les comptes liés à l'ambassade de Russie ont aussi été gelés pour *suspicion de blanchissement d'argent ou de financement de terrorisme*.

La gestion du risque est alors un processus continu, devant s'adapter aux contraintes du business, ou bien au micro-environnement de l'entreprise.

2.4.2 Structure

Pour s'aligner sur le budget fourni par la société mère, il faut non seulement comprendre le profil de risque actuel, mais aussi savoir le prédire. On parle donc de **gestion proactive** des risques. Il s'agit d'une des réponses possibles au risque, en plus de la réduction. C'est pour cela que l'on retrouve deux équipes dans la supervision des risques financiers :

- **Risk Control & Oversight** : pour le suivi du risque "actuel". Afin de faire face aux pertes liées et maintenir la stabilité financière, il faut établir les provisions adéquates. Elle collabore également avec l'équipe des risques non financiers en surveillant le risque opérationnel, notamment la Fraude.
- **Statistiques et Modélisation** : équipe pour **l'anticipation et la prévisibilité** du risque crédit. C'est au sein de cette équipe que c'est déroulé mon alternance.

3 La gestion du risque de crédit

Le risque de crédit est le principal risque financier auquel SC Banque est exposée. De manière générale, il se définit comme *le risque pouvant survenir à la suite d'un manquement aux obligations contractuelles convenues lors d'opérations financières*.

Cependant, il ne faut pas le confondre avec le risque d'octroi, car ils appartiennent à deux étapes différentes du processus de prêt. Le risque d'octroi est lié à l'évaluation initiale de la solvabilité des emprunteurs pour leur accorder ou leur refuser le crédit, c'est ce que les analystes Retail et Wholesale font chez SCB. En revanche, le risque de crédit est lié à la performance des contrats déjà financés et qui tombent en défaut. Celui-ci doit être analysé. C'est ce que l'on fait en tant que Credit Risk Analyst.

3.1 Types de risques de crédit

En réalité, le risque de crédit peut se décomposer en trois types de risques :

- **Risque de concentration** : conséquence de l'exposition à une seule contrepartie ou à un seul secteur. Comme le portefeuille de SCB est assez diversifié, ce risque ne représente pas une menace.
- **Risque de pays** : lié à l'instabilité politique et du pays de l'entreprise ou des partenaires. À ce stade, ce risque n'est pas pris en compte non plus puisque l'entreprise exerce son activité seulement en France métropolitaine.
- **Risque de défaut** : c'est le risque définit précédemment, et sur lequel porte le reste de ce rapport.

3.2 Définition du défaut

La définition précise du défaut varie en fonction du pays et des réglementations en vigueur. En France, celle-ci n'est pas restée figée au fil du temps. Avant la crise financière de 2008, un client en défaut était tout simplement un client en retard de paiement de plus de 90 jours. Cependant, après la crise financière, de nouvelles normes ont été introduites pour mieux refléter les risques réels associés aux prêts.

Pour information, la norme IFRS-9 (Norme internationale d'information financière) a été adoptée en 2018 en France, remplaçant la norme IAS-39. Cette norme a introduit une nouvelle approche pour la classification et l'évaluation des instruments financiers, y compris les prêts, mais aussi une nouvelle définition du défaut. Désormais, le défaut est défini comme la présence d'un *"événement de défaut"*, Ceci peut donc être un retard de paiement significatif, la faillite, etc.

Chez Santander Consumer Banque, la définition dépend du type de clientèle. Il faut savoir que la gestion du portefeuille chez SCB est scindée en deux modèles :

- la gestion **"standard" (Retail)** : clients particuliers et entreprises avec un encours < 100 000 euros et moins de trois véhicules.
- la gestion **"non standard" (Wholesale)** : entreprises ne vérifiant pas le critère cité ci-dessus et les concessionnaires.

Les critères de défaillance pour les deux segments sont assez similaires. J'ai alors choisi de définir le défaut pour les clients particuliers : **Un contrat en défaut est un contrat qui dépasse 90 jours de retard à tout moment au cours des 12 mois suivant la demande**, sinon il est considéré comme *"healthy"*. Le passage de l'état *Bad* à l'état *Good*, c'est-à-dire la détérioration de la solvabilité du client, se fait en trois étapes ou *stages*.

- **Stage 1** : Catégorie dans laquelle on retrouve les prêts présentant un faible risque de défaut. Les impayés de moins de 30 jours y compris.
- **Stage 2** : Catégorie concernant les prêts montrant des signes de détérioration de la qualité du crédit. Le passage en stage 2 se fait lorsque le client dépasse la date d'échéance de plus de 30

jours. On y trouve aussi les contrats en "*cure period*", c'est-à-dire les contrats ayant été classifié à un moment donné en stage 3, mais ils n'en font plus partie à l'heure actuelle. Les contrats de cette catégorie subissent des surveillances régulières.

- **Stage 3** : Catégorie concernant les prêts qui sont considérés comme étant en défaut de paiement. Il peut s'agir soit d'un client avec plus de 90 jours d'impayé, soit d'un client est en surendettement d'après la banque de France. À ce stade, des mesures de recouvrement ou de recouvrement des créances peuvent être mises en place pour récupérer les montants impayés.

Il faut noter que tout contrat s'avérant frauduleux sans relever d'un risque opérationnel est aussi classé en stage 3. Autrement dit, en fonction du motif de l'impayé, celui-ci peut être interprété ou pas comme un risque opérationnel. **La fraude peut donc aussi bien rentrer dans le risque de crédit que dans le risque opérationnel.** Exemple : une réjection de paiement due à des ressources insuffisantes dans le compte bancaire du client, peut suggérer qu'une activité frauduleuse est en cours (falsification des relevés bancaires, etc) mais il peut s'agir tout simplement d'un individu en difficulté financière.

3.3 Le processus de gestion du crédit de risque

L'évaluation des risques, dont le risque de crédit, est axée sur des mesures préventives. Après avoir fixé les objectifs et planifié le budget, la gestion du risque de crédit au sein de SCB suit le processus suivant.

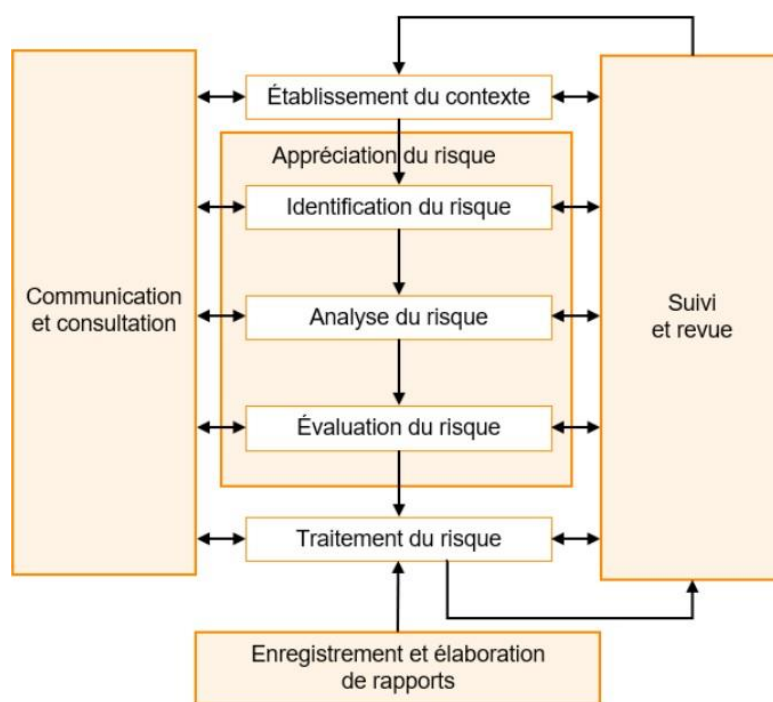


Figure 6 – Le processus de gestion du risque de crédit

J'ai principalement participé à la partie **appréciation du risque**.

1. **Identifier le risque** : Il s'agit ici d'identifier les "menaces", donc les demandes susceptibles de tomber en défaut. Cette phase représente le point d'entrée du risque en portefeuille. C'est pour cela qu'elle est encadrée par des outils de décision. En effet, les analystes suivent un processus décisionnel bien défini, constitué en premier temps un **Système de Notation**, plus couramment

appelé le **Scoring** puis d'un ensemble de règles politiques. **En fin de compte, les demandes sont : acceptés automatiquement, refusés automatiquement ou passent à l'analyse manuelle.** Actuellement, SCB dispose d'une seule "fiche de notation" ou *Scorecard* pour les clients particuliers :

— **Scorecard d'octroi** pour les automobiles

En conséquence, tous les dossiers SME, s'ils ne sont pas rejetés automatiquement, subissent une analyse manuelle.

Le Scoring d'octroi attribue à chaque demande **une note allant de 0 à 750 évaluant sa solvabilité**. Il est alors classifié dans une des quatre catégories : **BAD, AVERAGE, GOOD, VERY GOOD**. La politique de crédit diffère en fonction de la classe dans laquelle atterrit la demande.

Plus le Score est élevé, meilleur est le client, et moins il a de chance de tomber en défaut. C'est ce qu'on appelle la **Probabilité de Défaut : PD**. Elle estime la probabilité que l'emprunteur rembourse son prêt dans les termes du contrat. Elle est calculée à partir du Score. En conséquence, plus le score est élevé, plus la PD est faible, et inversement.

$$PD_{B2B} = \frac{1}{1+16 \cdot 2^{\frac{SCORE-580}{20}}} \quad (1)$$

$$PD_{B2C} = \frac{1}{1+16 \cdot 2^{\frac{SCORE-560}{20}}}$$

Le Score est calculé grâce à un **modèle statistique de régression logistique**. En fonction des caractéristiques du demandeur telles que sa situation professionnelle et financière, ses dettes, la durée du contrat, les caractéristiques du véhicule, etc. Pour des raisons de confidentialité, la liste des variables utilisés et leurs seuils ne seront pas donnés. Ces variables sont pondérées en fonction de leur importance. Par exemple, le **reste à vivre** et le **taux d'endettement** de l'individu pèsent beaucoup dans ce Score.

En tant qu'analyste de risque chez SCB, on ne construit pas nous-même le modèle car Santander Analytics (division du groupe en charge des modèles) l'achète directement chez **Experian**, une société mondiale de services d'information qui fournit des solutions d'évaluation du risque de crédit aux institutions financières. Cependant, il est primordial pour nous de bien **comprendre le fonctionnement cette première barrière de sécurité**. Ce fut l'objectif de mes études exploratoires.

2. **Analyser le risque** : Quantifier le risque de crédit à l'aide d'indicateurs **KPI : Key Performance Indicators** pour étudier le comportement du portefeuille d'une manière granulaire. Chaque mois, l'analyste de risque de crédit examine ces indicateurs afin détecter des anomalies, c'est-à-dire des comportements anormaux du risque. Pour s'y faire, on dispose d'un **outil de Reporting** permettant de visualiser l'évolution des KPI au fil du temps. Il existe plusieurs techniques d'analyses, mais chez SCB, on a souvent recours à ces deux-là :

- **Analyse comparative** : L'approche la plus explicite est de comparer les chiffres du mois actuel avec les données historiques recueillis lors des analyses précédentes.
 - **Analyse avec Forage de données** : Souvent, une analyse plus approfondie est requise, où il faut comparer le comportement d'un segment en particulier contre le reste du portefeuille. On essaie donc de trouver une corrélation entre le comportement observé et une ou plusieurs caractéristiques de la clientèle. Par exemple, si l'on remarque une hausse anormale du risque pour les prêts MOTOS chez les particuliers, il faut faire un *drill-through*, c'est-à-dire percer à travers les différentes variables pour expliquer cette anomalie. Il peut s'agir d'un seul prêt très volumineux qui pèse beaucoup dans la production du mois et qui s'est avéré frauduleux.
- C'est bien l'objectif principal du stage : **Construire un reporting mensuel automatisé, dynamique, guidant la réflexion menée par les analystes du risque de crédit.**

3. **Évaluer le risque** : Une fois le risque est quantifié, il faut essentiellement évaluer deux choses. En premier temps, l'acceptabilité des risques. Comme énoncé précédemment, on ne pourra pas éliminer l'intégralité des clients risqués (Risque résiduel). Il faut donc émettre un jugement sur la situation :

- **La fréquence de l'évènement** : Si la hausse du risque ce mois-ci est liée par exemple à une exposition de véhicules de loisirs lors de laquelle SCB a présenté ses offres de financement et a reçu un nombre important de demandes. Il s'agit d'un évènement qui a lieu qu'une seule fois par an. Il ne serait pas sage de bouleverser toute la stratégie d'acceptation pour éviter ce risque.
- **L'impact de l'évènement** : Si l'on conclut que cette hausse est due par exemple à une demande de financement d'une Porsche Cayenne, d'un montant très important (> 100K euros), et que pour une raison ou une autre le paiement des échéances a été rejeté, on ne pourra pas modifier la stratégie pour cette raison non plus.

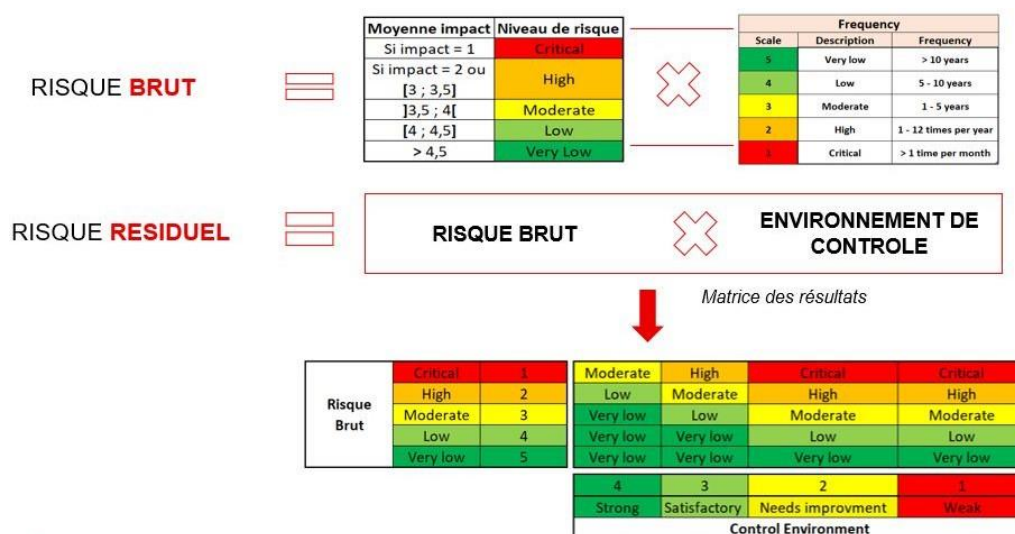


Figure 7 – Méthodologie pour le calcul du risque résiduel chez SCB

Pour faire face au risque, on a souvent recours à des **SDS Changes**, c'est à dire à modifier la stratégie d'admission. Avant toute chose, il faut évaluer les conséquences de l'action envisagée, surtout si on s'engage dans un changement profond qui impactera le fonctionnement de plusieurs équipes. C'est-à-dire tenir compte à la fois des effets immédiats, mais aussi des conséquences secondaires sur le système. En conséquence, une des missions est de **produire à partir des données des demandes déjà acceptées, des scénarios de et d'analyser leur impact.**

D'un point de vue modélisation, les possibilités de changement sont limitées. En effet, SCB n'intervient pas dans la construction du modèle, ni dans la définition des variables contribuant à la notation. Par contre, nous avons la main sur la calibration de certains seuils de discrétisation et surtout sur **l'ajustement de la valeur de Cut-off**. Cette valeur désigne **le seuil utilisé pour prendre la première décision**. En fonction de son Score, la demande peut être acceptée ou rejetée **automatiquement**, sans même besoin de l'étudier manuellement.



Figure 8 – Scoring : 1^{ère} étape de la décision de crédit chez SCB

Les Scénarios établis consistent à décaler le seuil de Cut-off de différentes manières et d'étudier l'impact de chaque décalage sur :

- **Le taux d'acceptation** : assouplir l'acceptation signifie plus de demandes. Ceci peut être le but, notamment dans le cadre d'un nouveau partenariat, ou pour attirer un nouveau type de clientèle. Sur le long terme, ceci pourrait nouer à la santé de l'entreprise.
 - **Le taux d'auto-décision** : Si l'on durcit le process d'admission automatique pour certains individus pour baisser le risque de défaut, cela augmenterait le nombre de décisions manuelles, et donc la charge de travail des analystes et le temps de décision.
 - **BPO** : Si l'on souhaite mettre en place des vérifications supplémentaires (documents, relevés bancaires, etc) pour baisser le risque de fraude, comment cela impacterait le coût opérationnel? Est-ce l'impact plausible ?
 - **Expected Loss** : les pertes potentielles liées à l'attitude que l'on décide de prendre : **conservatrice** (en étant plus prudent vis-à-vis des nouveaux clients) ou **agressive** (prendre du risque maintenant, mais maximiser les opportunités et augmenter le rendement sur le long terme.)
4. **Traitement du risque** : Une fois les scénarios présentés, le manager collabore avec les membres du directoire pour piloter le business en fonction des besoins et des volumes que l'entreprise reçoit actuellement (accélérer le process de décision, alléger la charge de travail des analystes, réduire le coût opérationnel, attirer une nouvelle clientèle, fidéliser une clientèle déjà existante, etc). En se basant sur le montant de risque évalué, on peut pratiquer le **Risk-Based Pricing**. On peut décider de facturer un taux d'intérêt plus élevé à la clientèle présentant un niveau de risque de défaut plus élevé. Plus concrètement, SCB propose des taux d'intérêts plus faibles à la clientèle TESLA. Il s'agit tout d'abord d'une opportunité commerciale (notoriété de la marque, ouverture précoce sur le marché des voitures électriques) mais il s'agit en plus d'une clientèle premium, avec des revenus plus élevés que la moyenne (67% des clients ont un revenu de plus de 2K euros/mois), donc moins risquée.

3.4 Situation Actuelle

Le processus d'octroi et le risque de crédit font l'objet d'une surveillance accrue dans les comités de contrôle des risques **ERCC**. Malgré le grand succès du nouveau score pour les particuliers développés en décembre 2020 (coût du risque de crédit baissé de 2 points et PD stabilisée), plusieurs enjeux restent encore sur la table.

En vue de la hausse du nombre de demandes en mars 2023 (+21%, soit près de 3k dossiers en plus en 1 mois), la durée de l'étude manuelle des dossiers se rallongeait de plus en plus. Le temps de décision moyen, aussi appelé **Time To Yes (TTY)** s'élevait à 36 heures pour les dossiers B2B. Aujourd'hui, le TTY, en comparaison à mars 2023, a drastiquement baissé et ces stabilisés autour de 14 heures. Pour ma part, il va falloir inclure cet indicateur dans mon reporting et en tirer des informations pour améliorer le processus.

La vente B2C constituée représente environ 30% du montant du New business enregistré chaque mois. En revanche, les produits DIRECT (prêts personnels) commercialisés n'ont pour l'instant rapporté que 1% à l'encours, puisque 80% des demandes ont été rejetées automatiquement. La direction des risques envisage alors **un assouplissement dans le processus d'admission en ce qui concerne ces demandes.**

Quant aux entreprises, malgré le fait qu'elles ne représentent que 20% du portefeuille, 40% d'entre elles montrent des signes de délinquance. Ces **NPL (Non-Performant Loans)** peuvent résulter de divers facteurs, tels que les séquelles de la pandémie de la Covid-19, ou bien l'inflation des prix et donc la baisse du reste à vivre pour les clients. Il faut donc continuer à **surveiller de près le comportement des SME**. Enfin, la fraude est un sujet très important à l'heure actuelle chez SCB. Le taux de fraudes avérées est en augmentation, avec des pertes majeures qui auraient pu être évitées. En effet, 12% des fraudes détectées proviennent d'une vingtaine de contrats financés. Des efforts sont en cours pour analyser la répartition motifs de rejets de paiements post-financement.

4 Description des missions

4.1 Déroulement

La première moitié de mon alternance fut consacrée à la phase exploratoire, à la construction et à l'automatisation des Reportings et aux analyses liées au système décisionnel.

La phase exploratoire consistait à explorer les notions présentées dans les deux chapitres précédents : découverte de l'entreprise, l'appréhension du fonctionnement, la compréhension du risque de crédit, etc. Au début, j'ai été confronté à un vocabulaire financier qui m'était totalement étranger. Les termes employés dans l'entreprise me semblaient complexes. Cependant, j'ai décidé de leur accorder assez de temps pour pouvoir pleinement m'immerger dans mon rôle par la suite. J'ai également profité des ressources internes mises à ma disposition, comme les documentations rédigées par la société mère et les formations obligatoires à faire en ligne. Pour comprendre les mécanismes de l'octroi, j'ai effectué quelques exercices sollicitant mon esprit logique. Cet apprentissage s'est fait graduellement, avec chaque nouvelle tâche que l'on me confiait.

Tout au long de cette expérience, j'ai travaillé en binôme avec mon tuteur qui m'a impliqué dans tous ses projets. Au début, nous collaborions afin de me guider. Petit à petit, je suis devenue plus autonome et capable d'exécuter les tâches par moi-même. Cependant, je m'adresse toujours à lui afin de vérifier la cohérence des résultats obtenus, puisqu'il connaît bien la composition et le comportement habituel du risque chez SCB.

4.2 Outils

C'était aussi l'occasion de me familiariser avec les outils utilisés pour établir nos reportings. Voici les outils qu'un Analyste des données risque chez SCB utilise au quotidien :

- **RAS (Risk Analytic Service)** : Il s'agit de l'outil d'informations End user. C'est une base de données relationnelle produite par **ORACLE**. Le RAS offre un environnement robuste et sécurisé permettant de : stocker, traiter et exporter une large quantité de données en peu de temps. Pour cela, on utilise le langage **SQL** dans des **requêtes** saisies et exécutées dans l'**IDE SQL Developer**.

RAS : le Datamart pour les analyses et la surveillance du risque. On y trouve toutes les données recueillies par SCB (dossiers, transactions, contrats, clients et actifs) provenant de différentes applications commerciales : Front et Back office. La plupart des données sont alimentées automatiquement quotidiennement depuis le centre d'information, que nous recevons le lendemain. Certaines informations liées au risque sont fournies par des solutions commerciales externes. Exemple : ARGUS pour les valeurs prévisionnelles des véhicules, FIBEN de la banque de France pour les cotations des entreprises, etc.) (Annexe IV)

Il existe essentiellement quatre point d'entrée de données :

- **Cassiopae** : Un système de gestion Front-to-Back utilisé par SCB au quotidien, de la génération des budgets jusqu'à l'approbation du financement et la gestion des contrats. Il fournit les données liées à la gestion des contrats.
- **NeoAuto** : Plateforme B2C développée par SCB où les clients peuvent simuler différentes options de financement pour l'achat d'un véhicule disponible sur le site. On recueille les informations saisies par les clients lors de la soumission de la demande de financement.
- **Experian** : Comme dit précédemment, Experian fournit les Scores et donc les décisions préliminaires basées sur cette note (avant les règles politiques). Il fournit aussi les informations des clients (situation financière et professionnelle, montant du prêt, prix de l'actif, etc.)
- **R** : Il s'agit d'un langage de programmation et d'un logiciel permettant de manipuler les données et d'en faire des statistiques. C'est un outil que j'avais déjà rencontré pendant mon

cursus académique. J'ai souvent eu recours aux bibliothèques suivantes : **DPLYR** (traitement des tables), **ODCB et DBI** (connexion à la base de données Oracle), **ggplot2** (data visualization) et **lubridate** pour la transformation et gestion des dates.

- **QuickBase** : Outil de pilotage Commercial. On y trouve divers Reportings contenant des informations sur les dossiers, les commerciaux, les concessionnaires, etc. Je m'en suis servie pour extraire les données nécessaires à la *Data Quality Checks* (voir section suivante), et pour m'informer sur le temps de décision (TTY) de certains dossiers.
- **Cassiopae** : Système utilisé par les analystes. Je m'en suis servie pour vérifier la cohérence entre les décisions prises par les analystes, et les décisions communiquées dans notre base de données (recherche par code du dossier ou par code acteur). Sachant que le RAS n'enregistre que la dernière décision, cet outil est utile lorsqu'on a besoin de voir l'intégralité du process d'admission. Par exemple, s'il y a eu un rappel de Score suite des modifications dans le dossier (appelées **Représentations**).
- **Package Microsoft Office 365** : J'ai utilisé **Excel** au quotidien pour mener mes analyses : construction des tableaux de pivots, manipulation des données exportées, suppression des doublons, filtrage des données, etc. La majorité des études ponctuelles qui m'ont été demandés devaient être rendus sous un format **xlsx**. J'ai utilisé **PowerPoint** quelques fois pour synthétiser et présenter nos besoins en termes de données au responsable du RAS. Enfin, **Power BI**, un outil de Business Intelligence d'agrégation, d'analyse et de visualisation des données, est la plateforme sur laquelle j'ai hébergé le Reporting Final.

4.3 Construction et Automatisation des Reportings

Le reporting des risques figure parmi les tâches centrales d'un Data Analyst de risque. Depuis 2020, la société mère SCF a renforcé cet aspect en mettant en place des obligations de reportage dans chaque département de SCB. Ce projet a requis beaucoup de ressources, mais permet d'évaluer en temps réel l'exposition au risque, ce qui facilite la prise des décisions nécessaires pour l'atténuer. C'est la première tâche qui m'a été attribué : **s'occuper des Reportings périodiques du risque de crédit** : de l'extraction et du traitement des données jusqu'à la production et l'envoi aux personnes concernées :

- **Le manager de l'équipe** : Des réunions de Risk Monitoring sont organisées tous les mois par le manager pour surveiller le comportement du portefeuille. Durant cette réunion, il faut présenter les nouveaux constats et les résultats de toute nouvelle étude menée. De plus, selon les besoins, j'ai été sollicitée pour faire des analyses ponctuelles que j'envoyais seulement au manager, après validation par mon tuteur.
- **Santander Consumer Finance HQ** : En plus d'un **MIS Management Information System** mensuel, il faut répondre à toute demande d'informations concernant les profils de risques, le statut des impayés, etc.

Reporting	Recipient	Frequency	Type of processing	Description	Application	Responsible	Alternate Responsible
MIS (report HQ)	SCF – Credit Risk	Monthly	Semi-Automatic	Report on main metrics related to the Retail application process (submissions, automation, override, decision and risk)	Qlikview RAS	J. Priou	R. Milani
Rapports Power BI	SCF – Risk monitoring	Monthly	Semi-Automatic	Range of Power BI reports used to monitor various aspects of Retail credit risk including : - Vintage risk monitoring (30+ and 60+ at 3m, 6m, 12m, FPD) - Acceptance process (acceptation rate by scoring, segment, sales channel etc.) - SME scoring output	Quickbase RAS	J. Priou	R. Milani

Figure 9 – Reportings du risque de crédit chez SCB

4.3.1 Data Quality Checks

Avant toute analyse, il faut s'assurer de la bonne qualité des données. Il faut savoir que celles-ci sont d'abord nettoyées et transformées puis alimentées dans le RAS. La plupart des données concernant les demandes sont alimentées automatiquement et quotidiennement depuis le centre d'information. Pour les demandes saisies le jour J, nous recevons leurs informations (exploitables) le J+1.

En outre, les tables portant sur les contrats et les paramètres de risques sont en général mises à jour à la fin de chaque mois (on parle de **clôture des données**). Il faut donc attendre trois jours ouvrés avant de pouvoir exploiter ces données. Une fois la mise à jour est faite, il faut tout de même effectuer une série de tests pour confirmer la complétude et la continuité des données, exemple : Nombre de demandes soumises par mois, nombre de contrats activés par mois, montant de la production par mois. Si l'on se rend compte d'incohérences dans les chiffres du dernier mois (par exemple, production trop faible par rapport à la moyenne des productions précédentes), il faut relancer la mise à jour de la table.

J'avais donc à **produire et à présenter au manager de l'équipe, en début de chaque mois, un document résumant 9 tests** vérifiant :

- la qualité des principaux indicateurs.
- le pourcentage de variables manquantes.
- la continuité de la base de données RAS.
- la cohérence entre RAS et Quickbase.

Ce rapport est actuellement semi-automatique. C'est-à-dire que l'on a précédemment codé sur R les fonctions effectuant et exportant les résultats de la plupart test sur R, mais certaines fonctions sont à revoir. De plus, l'importation des données est à faire manuellement. Voici les étapes à suivre :

1. **Extraction des données du RAS** : Il s'agit de requêtes SQL assez simple permettant d'extraire les données les plus récentes telles quelles, sans traitement préalables. Puisqu'on extrait l'intégralité des données portant sur le dernier mois, la tâche devient chronophage. Le but est donc de trouver un moyen pour connecter les données au programme pour minimiser l'intervention humaine et donc les sources d'erreurs, et pour accélérer la tâche.
2. **Exécution du code R** : Suite à des changements dans la structure de la base de données, une partie du programme n'est plus à jour. Il faut donc l'adapter.
3. **Analyse du résultat** : Le but est que l'analyste intervienne seulement à cette étape, où il **évalue le nombre d'échecs et enquête sur la source du problème**. Il peut s'agir d'une table erronée, ou d'un cas problématique que l'on ne traite pas. Il faut y apporter une solution.

4.3.2 MIS

C'est un rapport mensuel à rendre le 5 jour ouvrable du mois au HQ de SCF. Ce rapport **consolide les principaux indicateurs de risque et d'octroi dans de différentes segmentations (B2B, B2C, Particuliers, Entreprises) sur un historique de 36 mois à compter de la date de clôture**. Il est construit à partir de deux sources d'informations différentes : un extrait de Quickbase(Business Object) et un autre extrait des bases de données Oracle dans RAS. On y trouve 120 métriques. On peut les catégoriser de la manière suivante :

- **L'octroi : en quantité et en montant** : Nombre de demandes valides reçues, approuvées, rejetées ou en attente, montant total des nouvelles demandes, quantité et montant des *Override Decisions*, c'est-à-dire les demandes sur lesquelles il y a eu une représentation qui a conduit à une nouvelle décision.
- **Variables des caractéristiques du produit** : Durée moyenne du contrat, taux d'intérêt moyen, rapport moyen du montant/revenue, rapport moyen des apports/revenue, etc.
- **Variables des caractéristiques de l'emprunteur** : Score moyen des demandes approuvées, score moyen des demandes rejetées, montant des apports (*Downpayment*),
- **Indicateurs de risque** : Probabilité de défautet, Expected Loss, nombre de jours impayés pour les demandes approuvées, rejetées (automatiquement ou manuellement) ou financées, etc.

La démarche à suivre est pratiquement la même que pour les Data Quality Checks. Contrairement à ceux-ci, le travail d'analyse post-exécution est beaucoup plus élaboré. Il faut savoir que la société mère attribut tous les ans une note à SCB, basée sur la qualité des reportings fait au cours de l'année. Ce rapport en fait partie. Il faut donc être particulièrement prudent et bien vérifier la consistance des résultats obtenus avec les rapports des mois précédents. L'analyse consiste à calculer les écarts entre le rapport précédent et le rapport actuel. Il est possible d'obtenir des écarts significatifs pour les mois les plus récents (m-1 à m-4), car il peut y avoir des modifications dans les statuts des contrats ou des épurations du stock. En revanche, on ne doit pas observer des écarts pour les mois m-15 à m-36. Si c'est le cas, il faut enquêter sur à nouveau sur la source de cette anomalie.

5 Analyses relatives aux SDS changes

5.1 Définitions préliminaires

Le terme **SDS Changes** fait référence aux changements dans le *Scorecard* (système de notation) utilisé pour évaluer le risque de crédit des emprunteurs, donc le défaut. Chez SCB, le défaut est défini comme suit : **Un contrat ayant dépassé les 90 jours de retard de paiement à tout moment dans les 12 mois suivant son activation.**

Le **Score** a précédemment été défini comme une note allant de 0 à 750, attribuée à chaque emprunteur selon sa capacité à rembourser son crédit. Plus le score est élevé, plus la probabilité de défaut (PD) est faible, et inversement.

$$PD_{B2B} = \frac{1}{1 + 16 \cdot 2^{\frac{SCORE - 580}{20}}} \quad (1)$$

$$PD_{B2C} = \frac{1}{1 + 16 \cdot 2^{\frac{SCORE - 560}{20}}}$$

Cette note résulte d'un modèle statistique de régression logistique construit à partir de plusieurs variables relatives aux :

- Caractéristiques de **l'emprunteur**. Voir le tableau ci-dessous.
- Caractéristiques de **l'offre de financement** : durée du contrat, taux d'intérêt, montant des échéances, etc.
- Caractéristiques du **véhicule** : état (neuf ou occasion), marque et modèle, puissance du moteur, kilométrage d'origine, valeur résiduelle, etc.

Les variables numériques sont discrétisées en plusieurs classes. Elles sont ensuite pondérées en fonction de leur importance. Par exemple, le revenu mensuel et le taux d'endettement de l'individu impactent fortement sa solvabilité. C'est pour cela que le coût du risque des banques est corrélé à la dégradation de la situation macroéconomique. C'est ce qu'on a pu observer avec la crise sanitaire, durant laquelle le coût du risque a triplé en France suite à une vague de défauts sur les prêts.

Pour des raisons de confidentialité, la liste exhaustive des variables utilisées et leurs seuils de découpage ne seront pas communiqués.

Exemples de facteurs pris en compte dans un Score	
Particuliers	Entreprises
Âge et nationalité	Statut et Ancienneté de l'entreprise
Catégorie socioprofessionnelle et ancienneté	Secteur d'activité et forme juridique
Région de résidence et type de logement	Région d'activité
Situation familiale et composition du foyer	Nombre d'employés et qualité du climat social
Revenus, reste à vivre, taux d'endettement, présence d'autres crédits	Chiffre d'affaires, dettes fiscales, etc

Table 1 – Exemple de variables d'un système de Scoring

Pour rappel, les modèles de Scoring utilisés par SCB ont été développés grâce au concours de Santander Analytics (division de SCF chargé des modèles) en collaboration avec l'entreprise **Experian**.

En fonction des informations et des analyses du portefeuille que nous leur fournissons, SCF définit les variables rentrant en jeu, leurs poids et leur discrétisation et/ou classification.

Cela dit, les SDS changes peuvent donc inclure des mises à jour des critères de notation, des ajustements de pondération des variables et d'autres modifications pour améliorer la précision du modèle. Ces modifications ont lieu régulièrement pour but d'aligner le processus d'admission aux besoins du business. En effet, l'assouplissement ou la contraction de la politique d'admission doit répondre à des enjeux :

- **Commerciaux** : répondre aux nouveaux besoins d'une clientèle fidèle, cibler un nouveau segment sur le marché, étendre la base de clients.
- **Opérationnels** : réduire ou augmenter la charge de travail des analystes et les coûts opérationnels des processus automatiques.
- **Économiques** : tenir compte de la volatilité économique et de son impact sur la solvabilité des clients et s'aligner sur l'appétence au risque défini par la société mère.
- **Réglementaires** : se conformer aux exigences réglementaires et aux nouvelles normes.

5.2 Rappel de la stratégie d'octroi

Il est utile de rappeler qu'actuellement, SCB utilise un Scorecard pour les particuliers :

- Scorecard d'**octroi pour l'automobile**. Ce modèle implémenté en décembre 2020 a permis d'automatiser le processus d'octroi à 55% : 37% d'acceptation système et 20 % de refus automatique). Selon les **comptes sociaux de SCB publiés en 2021**, il a démontré une meilleure segmentation en risque (baisse du coût du risque de 3 points pour les nouvelles demandes) et un taux d'acceptation globale satisfaisant d'environ 60%.

Une demande de financement passe par les trois étapes suivantes avant d'être décidée :

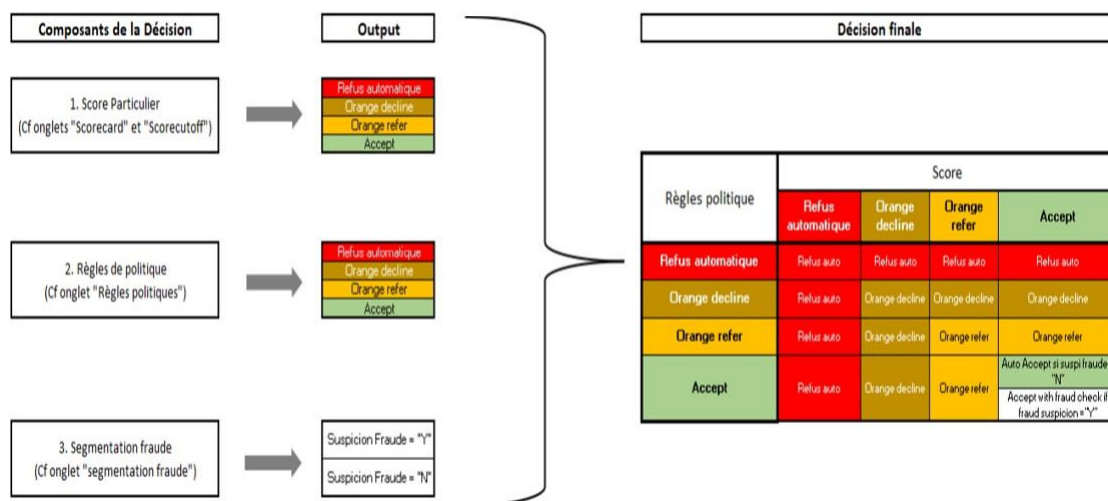


Figure 10 – Stratégie d'octroi pour les particuliers chez SCB

Si la demande n'est pas rejetée automatiquement par le système, elle passe obligatoirement par un deuxième filtre que sont les règles de politiques. Il existe essentiellement trois types de règles :

- **Règles de budget** : l'emprunteur ne doit pas dépasser un certain taux d'endettement (**Debt Ratio**). Il doit respecter un certain ratio endettements/revenus qu'on appelle le **Ratio Loan-to-Income**. Il existe aussi un minimum de **reste à vivre** à respecter. Le reste à vivre désigne la différence entre les revenus mensuels et les charges après le paiement des mensualités.
- **Règles de Blacklisting** : Fichage interne (concessionnaires radiés, individus ayant des antécédents de fraude, etc) ou externe (par la banque de France). Les demandes rentrant dans ces catégories sont refusées automatiquement.
- **Autres règles de SCB** : concernant la profession, l'âge, la durée du contrat, etc. Souvent, si les demandes ne respectent pas ces règles, l'analyste exige une modification des données (ajout d'un co-emprunteur, modification de la durée du contrat ou du montant des mensualités, etc). C'est ce qu'on appelle une **représentation**.

Enfin, toutes les demandes passent par la segmentation de Fraude. Les dossiers faisant l'objet d'une vigilance fraude sont intégralement étudiés à la main par l'équipe acceptation.

Une fois que le processus est mené, le statut du dossier est décrit par une variable appelée **Underwriting Level** :

- **Underwriting Level 1** : Refus automatique
- **Underwriting Level 4** : Acceptation automatique
- **Underwriting Level 2** : Revue manuelle
- **Underwriting Level 3** : Ok Suspicion Fraud

Il était impératif pour moi de comprendre tout ce cheminement, afin de pouvoir effectuer mes travaux d'analyse par la suite.

5.3 Changement de la Cut-off du Scoring Particulier

5.3.1 Contexte

Comme évoqué dans la partie précédente, les SDS changes ont lieu pour répondre aux besoins de la clientèle et pour aider à **mieux piloter** le business.

En effet, en mars 2023, notamment suite à la baisse des véhicules Tesla, la quantité de soumissions a augmenté de 25% par rapport aux quatre premiers mois de 2022. Face à cette hausse soudaine, les analystes de l'équipe d'acceptation se sont retrouvés à traiter un surplus de 3000 dossiers, dont 30% étaient des dossiers TESLA. En conséquent, le temps de décision, aussi appelé **Time-to-Yes** est passé de 15h à 36 heures pour les dossiers B2B. Sans compter la multiplication des aller-retours entre le département risque et le département des opérations situé en Pologne, couramment appelé **BPO** ou OPS.

Pour assurer une meilleure absorption de ce volume et pour anticiper les flux (véhicules de loisirs), il a fallu réfléchir à une nouvelle stratégie de Scoring plus automatisée. Avec plus de refus systématiques, on réduit le nombre de dossiers orientés en revue manuelle.

5.3.2 Analyse de scénarios

L'équipe de modélisation de SCB n'intervient pas dans la construction du modèle de Scoring, ni dans la définition des variables. En revanche, nous avons la possibilité d'apporter des changements dans la calibration du système (toujours en accord avec la société mère).

Pour ce type de besoin, nous avons souvent recours à un **réajustement de la valeur de Cut-off**, ou le seuil de coupure. Ce seuil numérique est utilisé pour décider si une demande est acceptée ou rejetée, ou passée en revue manuelle.

D'après le schéma précédent, il existe trois décisions pour le score : Accept (Vert), Refer, ou revue manuelle (Orange) et Refuse (Rouge). On a défini différents seuils de coupure pour différents segments du portefeuille, selon leur niveau de risque :

- Particuliers B2B avec un crédit <6k (**Moto**)
- Particuliers B2B avec un crédit >6k (**Auto et VDL**)
- **TESLA (B2C)**

Les dossiers Tesla ont plus de marge d'acceptation, car il s'agit non seulement d'un nouveau partenariat, mais aussi d'une clientèle "premium", responsable de 30% des apports mensuels, avec un risque de défaut quasiment nul (65% de la clientèle Tesla ont un revenu mensuel supérieur à 2000 euros).

La **Scenario Analysis** fait partie des tâches auxquelles j'ai participé. Je considère que cette phase a été essentielle à mon apprentissage. C'est à partir de ce moment-là que j'ai commencé à faire le lien entre les données, les processus d'octroi et le pilotage du business. Dans la gestion du risque, il faut avoir une vision ensembliste du risque, de l'acceptation et de la rentabilité.

Ce travail s'insère dans ce qu'on appelle la *Pré-prod*, c'est-à-dire tous les tests à effectuer avant de passer à l'action. Notre rôle est de calculer puis d'analyser l'impact du changement envisagé sur le système actuel. Le manager prend en compte toutes les répercussions potentielles avant de prendre sa décision. Pour s'y faire, nous avons considéré les mesures suivantes :

1. **Acceptante Rate (Taux d'acceptation)** : Chez SCB, le département du risque a adopté une approche **conservatrice**. On a défini un objectif d'**iso-acceptation**. La nouvelle valeur de cut-off ne doit pas faire passer plus ou moins de dossiers que d'habitude. Ce taux doit rester aux alentours des 60%.

$$\text{ACCEPTANCE RATE} = \frac{\text{Nombre de dossiers acceptés}}{\text{Nombre de dossiers décidés}} \quad (2)$$

2. **Auto-Decision Rate (Taux de décision automatique)** : Dans ce contexte, nous souhaitons l'augmenter le plus possible, tout en restant iso-acceptation.

$$\text{AUTO DECISION RATE} = \frac{\text{Nombre de demandes refusées ou acceptées automatiquement}}{\text{Nombre de demandes reçues}} \quad (3)$$

3. **BPO Impact** : c'est l'impact sur le coût opérationnel. Il faut savoir que toutes les demandes acceptées par le système passent obligatoirement chez BPO pour la vérification des documents de l'emprunteur via des logiciels spécialisés (pièces d'identités, relevés bancaires, attestations, etc). L'équipe BPO se charge aussi des opérations transactionnelles.
4. **Le risque** : pour évaluer l'impact sur la qualité globale du portefeuille, nous nous référons à l'indicateur **+30@MOB6**. C'est un des indicateurs que nous surveillons le plus tous les mois. Il est calculé à partir des échéances impayées reportées tous les mois. Il désigne le **pourcentage (en quantité et en valeur) de contrats ayant dépassé 30 jours d'impayés sur les six mois suivant leur activation**. Le terme MOB, abréviation de **Months-on-Books**, désigne le nombre de mois passés depuis le financement du dossier, autrement dit **l'ancienneté**. Il nous permet d'anticiper le risque de défaut. En 2024, le risque à court-terme, modélisé par cette mesure, s'est stabilisé.

Le +30@MOB6 se situe aujourd'hui aux alentours des 1%. Toute modification de la stratégie

d'octroi doit respecter ce seuil, et permettrait idéalement de le diminuer sur le long terme.

$$\begin{aligned}
 +30@MOB6_{Quantity} &= \frac{VIN_QUANTITY_DPD_30_6}{NEW_BUSINESS_QUANTITY} \\
 +30@MOB6_{Value} &= \frac{VIN_OUTSTANDING_DPD_30_6}{NEW_BUSINESS}
 \end{aligned}
 \tag{4}$$

avec :

- VIN_QUANTITY_DPD_30_6 : Quantité de contrats ayant un retard de paiement de plus de 30 jours sur leurs six premiers mois de vie. (DPD = *Days Past Due*).
- VIN_NEW_BUSINESS_DPD_30_6 : L'encours de ces derniers.
- NEW_BUSINESS_QUANTITY : La quantité de contrats activés ce mois-ci.
- NEW_BUSINESS : L'encours de ces derniers.

5. **Expected Loss** : Dans tous les cas, on cherchera toujours à minimiser les pertes attendues suite aux défauts. La direction impose un seuil maximal pour respecter le **Risk appetite** (appétence au risque).

$$EL = PD \cdot LGD \cdot EAD = PD \cdot \frac{0.588}{2}
 \tag{5}$$

avec :

- **PoD (Probability of Default)** : définie dans l'équation (1).
- **EAD (Exposure at Default)** : le montant total exposé par la banque au moment de la défaillance de l'emprunteur. En d'autres termes, c'est le montant du crédit à la date du défaut.
- **LGD (Loss-Given Default)** : la perte financière encourue par la banque en cas de défaut de paiement d'un emprunteur. Par exemple, si le LGD est de 58 %, cela signifie que la banque perdrait 58 % de l'EAD en cas de défaut.

(Se référer à l'annexe V pour les définitions et les formules des mesures du risque de crédit).

Pour trouver le seuil de coupure idéal, nous avons suivi un raisonnement empirique pour les trois segments (<6K, >6K et B2C). La majorité du travail est réalisée sous Excel. Voici un exemple de Scenario-Testing pour le segment des crédits B2B de montant > 6K euros :

1. **Extraire les données historiques** : Nous commençons par mettre en colonne toutes les valeurs possibles du Score, auxquelles nous associons les données de :
 - **Octroi** : le statut et le type de décision sont stockés dans la table **DATA_DEAL_DIMENSIONS** du RAS. On s'en sert pour les calculs (2) et (3)
 - **Risque** : la moyenne de la PD, la moyenne de la EL sont calculés à partir des formules (1) et (5) et du score stocké dans la table **RAT_DEAL_EXPERIAN_HISTORY**.
Les données pour le calcul (4) se trouvent dans la table **RDM_PORT_EVOLUTION_DET**.
 - **OPS** : nombre approximatif d'opérations par tranche de score.

Voici la première requête SQL que j'ai utilisée pour faire cette extraction. C'est l'un des premiers exercices que j'ai faits pour me familiariser avec la base des données, la structure des tables et surtout le langage SQL que je maîtrisais superficiellement auparavant.

Avec ces tables, il faut en général faire attention à prendre le dernier appel du Score, et le dernier mois de Reporting. Les trois tables sont reliées entre elles à l'aide de **jointures** basées sur le numéro du dossier, et sur le nombre de MOB. La clause **GROUP BY** segmente les données en tranches distinctes de Score, pour en suite y appliquer des agrégations (somme et division).

```

WITH decisions as (
select deal_id,
case when decision_last in ('Accepted', 'Refused') then 1 else 0 end as decided,
case when decision_last = 'Accepted' and if_last = '1' then 1 else 0 end as accepted
from data_deal_dimensions),

risk_data as (
select deal_id, contract_id, reporting_date, VIN_MONTHS_ON_BOOKS,
case when vin_months_on_books = '6' then 1 else 0 end as MOBE,

--- quantité de nouveaux contrats
VIN_NEW_BUSINESS,
VIN_NEW_BUSINESS_QUANTITY, --- montant des nouveaux contrats

VIN_QUANTITY_DPD30, --- nombre de fois où il y a eu un +30 jours retard de paiement
VIN_OUTSTANDING_DPD30 ---montant du +30 jours retard paiement

from RDM_PORT_EVOLUTION_DET)

select distinct
e.oscore,
(1/(1+16*power(2,((e.oscore - 580)/20)))) as PD,
(1/(1+16*power(2,((e.oscore - 580)/20))))*0.588/2 as EL,
sum(d.decided) as DECIDED_DEAL,
sum(d.accepted) as ACCEPTED_DEAL,
sum(d.decided)/count(distinct d.deal_id) as DECISION_RATE,
sum(d.accepted)/sum(d.decided) as ACC_RATE,

case when sum(r.VIN_NEW_BUSINESS_QUANTITY) = 0 then null else sum(r.VIN_QUANTITY_DPD30) / sum(r.VIN_NEW_BUSINESS_QUANTITY) end as plus_30_MOB_6_QTTY,
case when sum(r.VIN_NEW_BUSINESS) = 0 then null else sum(r.VIN_OUTSTANDING_DPD30) / sum(r.VIN_NEW_BUSINESS) end as plus_30_MOB_6_VAL

from PAT_DEAL_EXPERIAN_HISTORY e
left join decisions d on e.iapplicationnumber = d.deal_id
left join risk_data r on e.iapplicationnumber = r.deal_id and r.reporting_date = '31/03/2023' and r.MOBE = 1
group by oscore, decided, accepted, VIN_NEW_BUSINESS, VIN_NEW_BUSINESS_QUANTITY, VIN_QUANTITY_DPD30, VIN_OUTSTANDING_DPD30;

```

Figure 11 – Scenario-Testing : Requête SQL

2. **Définir les scénarios** : Pour chaque scénario, on choisit un seuil de coupure différent. Ce choix dépend du besoin. Ici, on voudrait augmenter le taux d'auto-décision (plus de lignes en rouge et en vert) :

- **Scénario 1 (Restrictif)** : +5 points d'auto-refus (Rouge) et +5 points d'auto-acceptation (Vert).
- **Scénario 2 (Libéral)** : +4 points d'auto-refus et +10 points d'auto-acceptation. Dans les deux cas on diminue les revues manuelles (Orange).

OSCORE	DECISION MAN.	ACCEPTED MAN.	REJECTED MAN.	MOB 6	SCENARIO 1	SCENARIO 2	Average of PD	Average of EL	+30 MOB 6 QTTY	POTENTIAL AUTO ACCEPT
120	0	0	0	0	0	0	0.00%	0.00%	0	0
121	0	0	0	0	0	0	0.00%	0.00%	0	0
122	0	0	0	0	0	0	0.00%	0.00%	0	0
123	0	0	0	0	0	0	0.00%	0.00%	0	0
124	0	0	0	0	0	0	0.00%	0.00%	0	0
125	0	0	0	0	0	0	0.00%	0.00%	0	0
126	0	0	0	0	0	0	0.00%	0.00%	0	0
127	0	0	0	0	0	0	0.00%	0.00%	0	0
128	0	0	0	0	0	0	0.00%	0.00%	0	0
129	0	0	0	0	0	0	0.00%	0.00%	0	0
130	0	0	0	0	0	0	0.00%	0.00%	0	0
131	0	0	0	0	0	0	0.00%	0.00%	0	0
132	0	0	0	0	0	0	0.00%	0.00%	0	0
133	0	0	0	0	0	0	0.00%	0.00%	0	0
134	0	0	0	0	0	0	0.00%	0.00%	0	0
135	0	0	0	0	0	0	0.00%	0.00%	0	0
136	0	0	0	0	0	0	0.00%	0.00%	0	0
137	0	0	0	0	0	0	0.00%	0.00%	0	0
138	0	0	0	0	0	0	0.00%	0.00%	0	0
139	0	0	0	0	0	0	0.00%	0.00%	0	0
140	0	0	0	0	0	0	0.00%	0.00%	0	0
141	0	0	0	0	0	0	0.00%	0.00%	0	0
142	0	0	0	0	0	0	0.00%	0.00%	0	0
143	0	0	0	0	0	0	0.00%	0.00%	0	0
144	0	0	0	0	0	0	0.00%	0.00%	0	0
145	0	0	0	0	0	0	0.00%	0.00%	0	0
146	0	0	0	0	0	0	0.00%	0.00%	0	0
147	0	0	0	0	0	0	0.00%	0.00%	0	0
148	0	0	0	0	0	0	0.00%	0.00%	0	0
149	0	0	0	0	0	0	0.00%	0.00%	0	0
150	0	0	0	0	0	0	0.00%	0.00%	0	0
151	0	0	0	0	0	0	0.00%	0.00%	0	0
152	0	0	0	0	0	0	0.00%	0.00%	0	0
153	0	0	0	0	0	0	0.00%	0.00%	0	0
154	0	0	0	0	0	0	0.00%	0.00%	0	0
155	0	0	0	0	0	0	0.00%	0.00%	0	0
156	0	0	0	0	0	0	0.00%	0.00%	0	0
157	0	0	0	0	0	0	0.00%	0.00%	0	0
158	0	0	0	0	0	0	0.00%	0.00%	0	0
159	0	0	0	0	0	0	0.00%	0.00%	0	0
160	0	0	0	0	0	0	0.00%	0.00%	0	0
161	0	0	0	0	0	0	0.00%	0.00%	0	0
162	0	0	0	0	0	0	0.00%	0.00%	0	0
163	0	0	0	0	0	0	0.00%	0.00%	0	0
164	0	0	0	0	0	0	0.00%	0.00%	0	0
165	0	0	0	0	0	0	0.00%	0.00%	0	0
166	0	0	0	0	0	0	0.00%	0.00%	0	0
167	0	0	0	0	0	0	0.00%	0.00%	0	0
168	0	0	0	0	0	0	0.00%	0.00%	0	0
169	0	0	0	0	0	0	0.00%	0.00%	0	0
170	0	0	0	0	0	0	0.00%	0.00%	0	0
171	0	0	0	0	0	0	0.00%	0.00%	0	0
172	0	0	0	0	0	0	0.00%	0.00%	0	0
173	0	0	0	0	0	0	0.00%	0.00%	0	0
174	0	0	0	0	0	0	0.00%	0.00%	0	0
175	0	0	0	0	0	0	0.00%	0.00%	0	0
176	0	0	0	0	0	0	0.00%	0.00%	0	0
177	0	0	0	0	0	0	0.00%	0.00%	0	0
178	0	0	0	0	0	0	0.00%	0.00%	0	0
179	0	0	0	0	0	0	0.00%	0.00%	0	0
180	0	0	0	0	0	0	0.00%	0.00%	0	0
181	0	0	0	0	0	0	0.00%	0.00%	0	0
182	0	0	0	0	0	0	0.00%	0.00%	0	0
183	0	0	0	0	0	0	0.00%	0.00%	0	0
184	0	0	0	0	0	0	0.00%	0.00%	0	0
185	0	0	0	0	0	0	0.00%	0.00%	0	0
186	0	0	0	0	0	0	0.00%	0.00%	0	0
187	0	0	0	0	0	0	0.00%	0.00%	0	0
188	0	0	0	0	0	0	0.00%	0.00%	0	0
189	0	0	0	0	0	0	0.00%	0.00%	0	0
190	0	0	0	0	0	0	0.00%	0.00%	0	0
191	0	0	0	0	0	0	0.00%	0.00%	0	0
192	0	0	0	0	0	0	0.00%	0.00%	0	0
193	0	0	0	0	0	0	0.00%	0.00%	0	0
194	0	0	0	0	0	0	0.00%	0.00%	0	0
195	0	0	0	0	0	0	0.00%	0.00%	0	0
196	0	0	0	0	0	0	0.00%	0.00%	0	0
197	0	0	0	0	0	0	0.00%	0.00%	0	0
198	0	0	0	0	0	0	0.00%	0.00%	0	0
199	0	0	0	0	0	0	0.00%	0.00%	0	0
200	0	0	0	0	0	0	0.00%	0.00%	0	0

Figure 12 – Scenario-Testing : Exemples de scénarios

3. **Estimer l'impact** : Pour chaque scénario, on calcule les mesures à l'aide de formules Excel. Ça a été l'occasion pour moi de revoir les formules basiques (SOMME, SOMMEPROD, SOMMESI, etc) et de m'entraîner sur les références relatives, absolues et mixtes. On résume les résultats dans un tableau :

		ISO-RANGES	ISO-ACCEPTATION
	ACTUAL	SCENARIO 1	SCENARIO 2
ACC RATE	59.61%	57.75%	58.85%
AUTO DECISION	61%	68.70%	69.53%
EL	1.14%	1.02%	1.07%
IMPACT BPO	-	5.35%	12.98%
+30@6	1.35%	1.30%	1.37%

Figure 13 – Scenario-Testing : Estimation des impacts

On voit que le deuxième scénario permettrait de maximiser le taux d'auto-décision, sans trop impacter le taux d'acceptation. En revanche, on s'attend à beaucoup plus d'efforts de la part de BPO. Le manager effectue le choix en fonction de la situation du budget et du recrutement.

5.3.3 Tests de régression et de non-régression

Une fois le choix fait, on code la nouvelle règle de cut-off dans le moteur de décision **Power Curve**. Je n'ai pas contribué à cette tâche, car elle requiert une formation préalable de la part de Experian. Avant de valider les changements et de passer à la *prod*, des tests de régression et de non-régression sont effectués pour voir l'impact sur les variables de notre base de données. Cette étape est cruciale, car une faille qui passe inaperçue peut engendrer des problèmes conséquents dans le système d'admission.

J'ai pu exécuter cette tâche en toute autonomie à l'aide de l'outil **R** et de sa librairie **Dplyr**. Ce fut un très bon exercice de compréhension du système d'octroi, de développement en R et surtout pour m'apprendre à synthétiser efficacement les informations. C'est à mon avis une des compétences les plus importantes à avoir pour ce métier.

Pour s'y faire, on se base sur deux échantillons de données extraits du moteur de décision :

- INPUT : demandes soumises **AVANT** la date d'implémentation du changement.
- OUTPUT : demandes soumises **APRÈS** la date d'implémentation du changement.

On s'intéresse aux variables relatives au Score et à la décision.

- **Tests de non-régression** : Il s'agit de confirmer que la nouvelle règle de cut-off n'introduit pas d'erreurs ou de déséquilibres dans les décisions d'approbation et de rejet par rapport aux performances antérieures du Scorecard.
- **Tests de régression** : Il s'agit de s'assurer que le changement se produit bien pour les demandes concernées par celui-ci, après la date d'implémentation.

J'ai procédé de manières similaires pour les trois changements. Continuons avec le cas du Scorecard B2B >6K. Après avoir importé les deux échantillons, on s'assure de prendre le dernier appel du Score et d'éviter les doublons en sélectionnant pour chaque dossier la ligne ayant la date de traitement la plus récente. (fonctions **arrange** et **slice(1)**). Ensuite, on isole le segment de clients concernés par le changement. (Particuliers HORS B2C ayant demandé un crédit >6k). On code ensuite les anciennes règles de segmentation dans l'échantillon de départ, et les nouvelles dans le deuxième.


```

#PLUS de 6K,PART & B2B
#CHECKS SCORE AUTODECISION
before_db<- read.csv("./csv/tasks/impact_score/data/output_NEUT.csv")
after_db <- read.csv("./csv/tasks/impact_score/data/output_vf_20230524.csv")

before_db <- before_db %>%mutate(INDEX = row_number())%>%
  arrange(Input.Application.Number,desc(INDEX)) %>%
  group_by(Input.Application.Number) %>%
  slice(1)%>%
  filter(Input.Loan.Amount >= 6000)%>%
  filter(Input.Client.Type == 'PART') %>%
  filter(substr(Input.Application.Number,1,5) != 'OFR20') %>% filter(Input.Dealer.ID!="A00473554")

after_db <- after_db %>%mutate(INDEX = row_number())%>%
  arrange(Input.Application.Number,desc(INDEX)) %>%
  group_by(Input.Application.Number) %>%
  slice(1)%>%
  filter(Input.Loan.Amount >= 6000)%>%
  filter(Input.Client.Type == 'PART') %>%
  filter(substr(Input.Application.Number,1,5) != 'OFR20')%>% filter(Input.Dealer.ID!="A00473554")

before_db_d <- before_db %>%mutate(SEG = case_when(Output.Score <= 570 & Output.Score >= 550 ~ 'ORANGE',
  Output.Score < 550 ~ 'ROUGE',
  Output.Score > 570 ~ 'VERT'))

after_db_d <- after_db %>%mutate(SEG = case_when(Output.Score < 561 & Output.Score >= 554 ~ 'ORANGE',
  Output.Score < 554 ~ 'ROUGE',
  Output.Score >= 561 ~ 'VERT'))

```

Dans les tests de régressions, on s'assure que les décisions non concernées par le changement gardent le même score, la même décision de score et le même underwriting level. Pour s'y faire, on les "flag" dans l'échantillon de départ puis on les retrouve grâce à une jointure dans le deuxième échantillon. On examine leurs variables grâce à la fonction **diffdf**, permettant de comparer deux datasets. On les examine aussi de plus près en exportant des cas dans Excel et en menant les vérifications à la main. (Ici, on introduit une nouvelle règle de politiques pour les clients déjà existants, en plus du changement de la cut-off que je traite aussi.).

```

#TEST NON REGRESSION 1: ORANGES RESTENT ORANGE
##OK
oranges_before <- before_db_d%>%filter(Output.Score < 561,Output.Score >= 554)%>%
  select(Input.Application.Number, INDEX,SEG, Output.Score, Output.Score.Decision)

check_1 <-oranges_before%>%
  inner_join(after_db_d%>%select(Input.Application.Number,INDEX, SEG, Output.Score, Output.Score.Decision),
    by= c("INDEX", "Input.Application.Number" ))

#TEST NON REGRESSION 2: NON AFFECTED DEALS STAY THE SAME

non_affected_before<- before_db_d%>%filter(Output.Score > 570 | Output.Score < 550)%>%
  select(Input.Application.Number, INDEX,SEG, Output.Score, Output.Score.Decision)

non_reg <-non_affected_before%>%
  inner_join(after_db_d%>%select(Input.Application.Number,INDEX, SEG, Output.Score, Output.Score.Decision),
    by= c("INDEX", "Input.Application.Number" ))

#variables ne changent pas pour les rouges/verts non concernés, les existing customers should not be in susp fraud
var_1<- before_db_d%>% filter(Input.Main.App.Existing.Customer=="N")%>%
  filter(Output.Score > 570 | Output.Score < 550)%>%inner_join(after_db_d%>% filter(Input.Main.App.Existing.Customer=="N"),
    by=c("INDEX", "Input.Application.Number"),
    suffix = c('_before', '_after'))

non_reg_var_1 <- diffdf(var_1[, endswith(colnames(var_1), "_before")]%>%
  rename_with(~str_remove(., '_before')),
  var_1[, endswith(colnames(var_1), "_after")]%>%
  rename_with(~str_remove(., '_after')))

```

Figure 14 – Tests de non-régression sur R

Le même raisonnement est utilisé dans les tests de régression. En revanche, ceux-ci sont en général plus problématiques et requièrent plus de réflexion, car il y a un tas de règles de politiques à garder en tête en plus de la décision du Score. En effet, lors de cette étape, nous nous sommes rendu compte de plusieurs failles dans le moteur de décision. Comme il a fallu répéter l'exercice à chaque fois, j'ai fini par repérer les cas problématiques à vérifier en priorité. Ce qui m'a facilité la tâche par la suite. (Se référer à l'Annexe VI pour un aperçu du code R). Ces travaux sont à reporter au manager de l'équipe, à documenter et à sauvegarder pour des fins de contrôle interne. J'ai alors constitué un "log" sur Excel, pour documenter de manière organisée toutes les étapes, les observations et les détails pertinents. (Annexe VI)

5.3.4 Calcul et analyse de l'impact

Nos travaux ont estimé une réduction entre 10% et 15% des dossiers revus manuellement. Une de mes missions est d'assurer un suivi des actions mises en place, pour confirmer si les objectifs ont bien été atteints. Il se peut qu'on ait des écarts entre nos estimations et la réalité, car nos scénarios se basent sur les données historiques des dossiers décidés.

Pour s'y faire, j'ai utilisé les mêmes outils et la même méthode de traitement des données : distinction des données avant et après implémentation, isolation du segment concerné, clauses group by pour calculer la répartition de la demande par type de décision, etc. Voici les résultats obtenus après le changement du Score Particulier pour le segment B2B.

TOTAL B2B						
Output.Underwriting.Level	BEFORE		AFTER AUTODECISION STRATEGY		IMPACT	
	# deals	% deals	# deals	% deals		
1	7783	17.21%	8923	19.73%	2.52%	
2	18733	41.42%	15459	34.18%	-7.24%	
3	4387	9.70%	5184	11.46%	1.76%	
4	14329	31.68%	15666	34.63%	2.96%	

Figure 15 – Score Particulier : Analyses de l'impact réel

On voit bien qu'on gagne +5.5% en auto-décision (level 4 et level 1) et qu'on perd 7% de revues manuelles, soit une baisse totale de 12.5% du flux pour l'équipe acceptation. C'est l'équivalent de la charge en dossiers de deux analystes. Donc l'objectif a été atteint du côté B2B.

6 Projets

Au début de mon alternance, je n'étais pas attribuée à un planning de tâches spécifiques. Cependant, l'essence même de cette expérience consistait à apporter mon soutien à mon tuteur dans divers projets en cours. Ainsi, deux projets ont suivi : la reconstruction d'un outil de Reporting pour le département des risques sur Power BI et la mise en place de code Sql et R pour la détection de Fraude.

6.1 Les besoins et mes propositions

6.2.1 Projet Fraude

Le but ultime de ce projet est de **créer des codes automatisés et facile à utiliser**, mis à disposition à l'ensemble du département des risques. L'intérêt étant d'avoir des listes d'individus suspectés de fraude en prenant en compte les critères communs aux fraudeurs. On avait décidé après réflexion sur les cas de fraude précédent, de lister les dossiers ayant un mail identique à un ou des dossiers antérieurs pour un même « name » (Prénom + Nom) ou pour un même identifiant de concession. Soit deux listes.

C'est pourquoi dans un premier temps, j'ai assemblé l'ensemble des variables des dossiers de crédit les plus pertinents pour détecter une fraude.

J'ai donc tiré les dossiers sur SQL afin de pouvoir les traiter sur R via la mise en place de codes.

A terme, on souhaiterait avoir des code R qui puisse nous éviter l'extraction sur SQL et de pouvoir tirer les listes directement avec les codes R. Mais aussi, de pouvoir tirer des listes de plus en plus simplifiées pertinentes.

6.2.2 Projet PBI

Le but ultime de ce projet est de **créer un outil de Reporting : complet, centralisé, automatisé et facile à utiliser**, mis à disposition à l'ensemble du département des risques. Idéalement, il comporterait un :

- 1. Module pour le **Risk Monitoring** : pour la surveillance des indicateurs du risque de crédit (sur une base mensuelle) et pour l'analyse du comportement du portefeuille.
- 2. Module pour les **analystes** : destiné aux analystes seniors de l'équipe d'acceptation (données relatives aux travaux des analystes, quantité et qualité des demandes traitées, temps de décision, etc).
- 3. Module de **Granting** : pour l'analyse de la qualité des demandes reçues, des profils des emprunteurs (PD, EL).
- 4. Module de **Contrôle interne** : destiné au manager de l'équipe acceptation. À utiliser une fois par mois pour la vérification du respect de la stratégie d'octroi.

D'un point de vue **User Experience**, on assurerait la navigation entre les modules grâce à des menus et des boutons intuitifs. On introduirait des fonctionnalités dynamiques pour l'analyse approfondie du comportement du risque.

Le processus nécessite quand même une intervention humaine pour l'exécution d'un script SQL, ensuite d'un programme R et enfin d'une requête Power Query.

Cette approche représente un pas significatif vers l'amélioration de la gestion du risque.

6.2 Production Projet Fraude

Une fois les objectifs définis, nous avons organisé des réunions pour décider des variables les plus intéressantes à mettre en avant en prenant en compte les précédents dossiers frauduleux. L'objectif étant de juger quelles sont les variables qui sont le plus enclin à être modifiées par le client et qui méritent une attention sur le dossier.

6.2.3 Collecte des données

La première étape de ce projet est la collecte des données. Ça commence par les données provenant du RAS.

Pour minimiser la durée d'extraction et la taille des données, j'ai limité les dossiers sur une durée de 1 an avant l'extraction et on liste pour chaque table les variables dont on a besoin.

À la suite de cette préparation, on code une **requête globale en SQL**.

```
SELECT a.transeg_bizkey, a.IAPPLICATIONNUMBER, c.DEAL_ID_FIRST, d.first_name, d.last_name,
       a.ILOANAMOUNT, a.IMA_NATIONALITY, UPPER(a.IMA_EMAILADDRESS) AS email_address, a.IAPPLICATIONDATE,
       a.idealid, ACTOR_ID_CO_CLIENT, actor_id, c.decision_last, a.OUNDERWRITINGLEVEL, c.REJECTION_REASON_CODE,
       a.IMA_ZIPCODE, c.TOTALINCOME, a.IMA_MARITALSTATUS, a.IMA_NUMBEROFCHILDREN, a.IMA_EXISTINGCUSTOMER, a.IEXISTINGCUSTOMER,
       a.ICLIENTTYPE, a.OREASONCODES_I1, a.OREASONCODES_I2, a.OREASONCODES_I3
FROM RAT_DEAL_EXPERIAN_HISTORY a
LEFT JOIN data_deal_dimensions c ON a.IAPPLICATIONNUMBER = c.DEAL_ID
LEFT JOIN data_actor_dimensions d ON c.ACTOR_ID_CLIENT = d.actor_id
WHERE a.IAPPLICATIONDATE >= TRUNC(SYSDATE) - INTERVAL '1' YEAR AND a.IAPPLICATIONDATE < TRUNC(SYSDATE);
```

6.2.4 Traitements des données

Une fois les données extraites et rassemblées, on procède au nettoyage et au traitement des données sur Rstudio. Par exemple pour les listes de suspicions de fraude basées sur des mail communs sur différents dossiers avec des 'name' différents :

Il s'agit de

- Convertir la colonne de date IAPPLICATIONDATE au format Date.
- Créer un tableau df avec des noms normalisés et la suppression de doublons
- Créer un tableau df1 avec des informations supplémentaires et calculer la date maximale.
- Analyser les doublons et identifier les incohérences dans les données
- Sauvegarder les données traitées dans un format facilement partageable.

```

Liste 1 Suspicion de fraude:
-----
library(dplyr)
library(stringr)

gb7 <- read.table("C:/Users/N271811/OneDrive - Santander Office 365/Desktop/gb7.dsv", header = TRUE, sep = ";")
gb7$IAPPLICATIONDATE <- as.Date(gb7$IAPPLICATIONDATE, format = "%d/%m/%y")

df <- gb7 %>%
  mutate(name = str_trim(toupper(paste(FIRST_NAME, LAST_NAME)))) %>%
  group_by(name, EMAIL_ADDRESS) %>%
  select(name, EMAIL_ADDRESS, IAPPLICATIONDATE) %>%
  distinct()

df1 <- gb7 %>%
  mutate(name = str_trim(toupper(paste(FIRST_NAME, LAST_NAME)))) %>%
  select(IAPPLICATIONNUMBER, DEAL_ID_FIRST, IDEALERID, name, EMAIL_ADDRESS, IAPPLICATIONDATE, ACTOR_ID, ACTOR_ID_CO_CLIENT, DECISION_LAST, OUNDERWRITINGLEVEL, REJECTION_REASON_CODE, ICLIENTTYPE, OREASONCODES_I1, OREASONCODES_I2, OREASONCODES_I3) %>%
  distinct() %>%
  group_by(name, EMAIL_ADDRESS) %>%
  mutate(max=max(IAPPLICATIONDATE))

df2 <- df %>%
  group_by(EMAIL_ADDRESS) %>%
  summarise(n = n(), unique_names = n_distinct(name)) %>%
  filter(n > 1 & unique_names > 1 & unique_names < 10) %>%
  left_join(df, by = "EMAIL_ADDRESS") %>%
  filter(IAPPLICATIONDATE >= as.Date("2024-08-07")) %>%
  left_join(df1, by = "EMAIL_ADDRESS") %>%
  distinct() %>%
  group_by(name.y, IAPPLICATIONDATE.y, EMAIL_ADDRESS) %>%
  arrange(IAPPLICATIONDATE.y) %>%
  slice(1)

df3 <- df2 %>%
  filter(n != unique_names) %>%
  arrange(EMAIL_ADDRESS)

df3 <- df3 %>%
  arrange(EMAIL_ADDRESS, name.y, desc(IAPPLICATIONDATE.y)) %>%
  group_by(EMAIL_ADDRESS, name.y) %>%
  slice(1)

df3 <- df3 %>%
  arrange(EMAIL_ADDRESS)

df3 <- df3 %>%
  select(IAPPLICATIONNUMBER, DEAL_ID_FIRST, EMAIL_ADDRESS, n, unique_names, name.y, IDEALERID, ACTOR_ID, ACTOR_ID_CO_CLIENT, max, DECISION_LAST, OUNDERWRITINGLEVEL, REJECTION_REASON_CODE, ICLIENTTYPE, OREASONCODES_I1, OREASONCODES_I2, OREASONCODES_I3)

df3 <- df3 %>%
  arrange(EMAIL_ADDRESS)

write.csv(df3, "C:/Users/N271811/OneDrive - Santander Office 365/Desktop/liste1_08aout.csv", row.names = FALSE)

```

6.2.5 Impact du projet de Fraude

Le projet a eu un impact significatif sur les opérations de l'équipe Risque et sur l'entreprise. Grâce aux analyses et aux algorithmes mis en place, de nombreux cas de fraude ont été détectés de manière proactive, permettant ainsi à l'entreprise de prévenir des pertes financières potentielles importantes. Cette amélioration à la détection a non seulement renforcé la sécurité des octrois, mais a également accru la confiance des clients envers nos systèmes.

Suite à la présentation des résultats après un mois de mise en place lors d'une réunion avec le PDG, le travail réalisé a été particulièrement salué. J'ai reçu des félicitations de la part du PDG et des manager, ainsi que des remerciements pour la qualité et l'efficacité de mon travail. Cette reconnaissance témoigne du projet et de sa contribution positive à l'amélioration des processus de gestion du risque au sein de l'entreprise.

6.3 Production Projet PBI

Une fois les objectifs définis, nous avons organisés des réunions pour noter les demandes en termes de données, de visuels, etc.

6.3.1 Collecte des données

La première étape de ce projet est la collecte des données. Ça commence par les données provenant du RAS. Pour minimiser la durée d'extraction et la taille des données, on liste pour chaque table les variables dont on a besoin. Ensuite, j'ai défini la manière dont elles sont liées (type de jointure et base de jointure).

Voici un extrait de ce Mapping.

REPORTING TYPE	TABLE	EXTERNAL FILES	JOINS
CREDIT RISK	RAT_DEAL_EXPERIAN_HISTORY DATA_DEAL_DIMENSIONS RDM_PORT_EVOLUTION_DET DATA_CONT_WRITEOFF_HISTORY	SME Scoring VBA REGION DEALER.xlsx SCORE FRAUD VARIABLES.xlsx	LEFT : IAPPLICATIONNUMBER = Deal_ID LEFT : IDEALERID = DEALER_ID LEFT : DEAL_ID = DEAL_ID RIGHT : CONTRACT_ID+REPORTING_DATE

Table 1 – Extrait d'un Mapping de données

À la suite de cette préparation, on code une **requête globale en SQL**.

```
LEFT JOIN PORT_PF ON (A.IAPPLICATIONNUMBER = TRIM(PF.DEAL_ID) AND PF.REPORTING_DATE = '31/07/2023')
LEFT JOIN PORT_SUB_1 G1 ON (A.IAPPLICATIONNUMBER = TRIM(G1.DEAL_ID_1) AND G1.VIN_MONTHS_ON_BOOKS_1 = 1 AND G1.REPORTING_DATE_1 <= '31/07/2023')
LEFT JOIN PORT_SUB_2 G2 ON (A.IAPPLICATIONNUMBER = TRIM(G2.DEAL_ID_2) AND G2.VIN_MONTHS_ON_BOOKS_2 = 2 AND G2.REPORTING_DATE_2 <= '31/07/2023')
LEFT JOIN PORT_SUB_3 F ON (A.IAPPLICATIONNUMBER = TRIM(F.DEAL_ID_3) AND F.VIN_MONTHS_ON_BOOKS_3 = 3 AND F.REPORTING_DATE_3 <= '31/07/2023')
LEFT JOIN PORT_SUB_6 D ON (A.IAPPLICATIONNUMBER = TRIM(D.DEAL_ID_6) AND D.VIN_MONTHS_ON_BOOKS_6 = 6 AND D.REPORTING_DATE_6 <= '31/07/2023')
LEFT JOIN PORT_SUB_9 Y ON (A.IAPPLICATIONNUMBER = TRIM(Y.DEAL_ID_9) AND Y.VIN_MONTHS_ON_BOOKS_9 = 9 AND Y.REPORTING_DATE_9 <= '31/07/2023')
LEFT JOIN PORT_SUB_12 E ON (A.IAPPLICATIONNUMBER = TRIM(E.DEAL_ID_12) AND E.VIN_MONTHS_ON_BOOKS_12 = 12 AND E.REPORTING_DATE_12 <= '31/07/2023')
```

Figure 16 – Aperçu de la requête globale (SQL)

Cette structure facilite aussi le **"Vintage Analysis"** ou l'analyse par cohorte. C'est une méthode d'évaluation de la performance du portefeuille au fil du temps. Le "Vintage Analysis par Months on Books" (ou MoB) est une approche spécifique qui évalue sur **la performance des prêts en fonction de la durée depuis leur origination**. C'est comme ça que l'indicateur +30@MOB6 présenté dans le chapitre I est calculé.

Une des nouveautés est le calcul du Time-To-Yes total, (du début de la soumission de la demande jusqu'à la signature du contrat), le TTY exclusif au département des risques, le TTY des revues manuelles et le **nombre d'aller-retours entre le département des risques et les OPS** pour chaque demande. Cette dernière information est très pertinente, car elle permet de quantifier les difficultés que rencontrent les OPS en ce moment, traduit par des durées de traitements plus importants et des échanges plus fréquents avec nos analystes.

```
TTY as (
  select deal_num,
    sum(case when duree_jalon is not NULL and date_fin is not null then DUREE_JALON/(60*60) else 0 end) as TTY_TOTAL
  from data_deal_jalon_timing_x
  group by deal_num),

TTY_RISK as (
  select deal_num,
    sum(case when jalmetier = 'RISK' and duree_jalon is not NULL and date_fin is not null then DUREE_JALON/(60*60) else 0 end) as TTY_RISK,
  from data_deal_jalon_timing_x
  group by deal_num),

TTY_MANUAL_STUDY as (
  select deal_num,
    sum(case when jalcode in ('SUNRISK', 'NOPEP', 'UNDPEP', 'PEPRISK', 'PEPOK')
      and duree_jalon is not NULL and date_fin is not null
      then DUREE_JALON/(60*60) else 0 end) as TTY_MANUAL_STUDY
  from data_deal_jalon_timing_x
  group by deal_num),

BACK_AND_FORTH as (
  select deal_num, sum(count_back_forth) as SUM_BACK_AND_FORTH_RISK_BPO from (
    SELECT deal_num,
      case when (jalmetier in 'RISK' and previous_jalmetier in 'BPO')
        then 1 else 0 end as count_back_forth
    FROM (
      SELECT deal_num, jalcode, jalmetier,
        LAG(jalmetier) OVER (PARTITION BY deal_id ORDER BY date_debut) AS previous_jalmetier
      FROM data_deal_jalon_timing_x
      WHERE jalmetier IN ('RISK', 'BPO')
    ) deal_transitions ) j
  group by deal_num),
```

Figure 17 – Méthode de calcul du TTY (SQL)

Pour le calcul des différents TTY, on se base sur **la somme des durées des jalons**. Un jalon est un code représentant une étape dans le traitement d'une demande.

$$\begin{aligned}
 TTY_{TOTAL} &= \sum_{i=1}^n \text{duree jalons} \\
 TTY_{RISK} &= \sum_{i=1}^n \text{duree jalons risque} \\
 TTY_{MANUAL_STUDY} &= \sum_{i=1}^n \text{duree jalons revue manuelles}
 \end{aligned}
 \tag{6}$$

avec :

- n : nombre d'étapes dans le traitement d'une demande
- durée jalon (h) : $\frac{\text{date de fin du jalon} - \text{date de début du jalon}}{3600}$

6.3.2 Traitements des données

Une fois les données extraites et rassemblées, on procède au nettoyage et au traitement des données dans un premier temps sur Rstudio, puis avec Power Query de Power BI. Il s'agit de

- Remplacer certaines valeurs manquantes par des 0 avant de les transformer en type numérique.
- Identifier et supprimer les valeurs extrêmes ou aberrantes
- Mettre toutes les variables de type date sous le même format pour pouvoir s'en servir correctement pour l'analyse temporelle du risque.
- Faire la jointure entre des fichiers externes et les données extraites via le RAS.
- Ajout de variables pour la mesure du risque : PD, **FPD_FLAG (variable binaire indiquant si l'emprunteur a eu un retard de paiement)**.
- Coder les règles du Score Fraude à la main pour en déduire la probabilité de défaut (nouveau).

```
#####
#DATA CLEANING & FORMATTING
#CONVERT COLUMNS TO NUMERIC
cols.num <- c('IMA_NETMONTHINCOME',
              'PERCENTAGEOFDOWNPAYMENT',
              'MONTHLYINSTALLMENT',
              'OMA_TIMEATCURRENTJOB',
              'IMA_MONTHINSTALLMENTHOMELANS',
              'IMA_MONTHINSTALLMENTOTHERLOANS',
              'IMA_NUMBEROFCHILDREN',
              'IAGEOF CAR',
              'OMA_AGE',

              'ORATIOLOANTOINCOME',
              'OTIMEATCURRENTADDRESS',
              'OMA_TIMEATBANK')

clean_numeric(global_extract[cols.num], remove = "[^0-9.]", fixed = FALSE)
global_extract[cols.num] <- sapply(global_extract[cols.num], as.numeric)

#DUE DATE and PAYMENT DATE are different date formats
global_extract$INVOICEDATE <- dmy(global_extract$INVOICEDATE)
global_extract$DUE DATE <- dmy(global_extract$DUE DATE)
global_extract$PAYMENT DATE <- dmy(global_extract$PAYMENT DATE)

date_cols <- grep("DATE", names(global_extract))
global_extract[, date_cols] <- lapply(global_extract[, date_cols], function(x) as.Date(x, format="%d/%m/%y", origin='1970-01-01'))

#INITIAL TREATMENTS
#replace NA with 0 and convert to numeric
list_num_vars <- grep("VIN", names(global_extract))
global_extract[, list_num_vars] <- sapply(global_extract[, list_num_vars], as.numeric)

global_extract[, list_num_vars] %>% replace(is.na(.), 0)
global_extract$INPL_BALANCE %>% replace(is.na(.), 0)

#PD FOR PART
#PD FOR SME
global_extract <- global_extract %>%
  left_join(DEALS_SME_RATING1, by = c('APPLICATIONNUMBER' = "DEAL_ID" ))

global_extract <- global_extract %>%
  left_join(REGION_DEALER, by = c("IDEALERID" = "DEALER_ID"))%>%
  left_join(REGION_DEALER_DETAILED, by = "IDEALERID")%>%
  left_join(DEALER_RATING, by = c("IDEALERID" = "IDEALER_ID"))%>%
  mutate(NEW_CLASSIFICATION = case_when(is.na(NEW_CLASSIFICATION) ~ 'SMALL',
    TRUE ~ NEW_CLASSIFICATION))%>%

  |
  mutate(GRP_DEALER_SCR = case_when( is.na(NEW_CLASSIFICATION) ~ 61,
    NEW_CLASSIFICATION == 'SMALL' ~ 61,
    NEW_CLASSIFICATION == 'MIDDLE - BIG' ~ 64,
    NEW_CLASSIFICATION == 'VERY BIG' ~ 84),

    SCORE_CALC = IPURPOSEOFLOAN_SCR + GRP_DEALER_SCR + OMA_TIMEATBANK_SCR + ITYPEOF CAR_SCR + OTIMEATCURRENTADDRESS_SCR + ORATIOLOANTOINCOME_SCR
    + MARITAL_BY_NBCHILDREN_SCR + PERCENTAGEOFDOWNPAYMENT_SCR + TIMEATJOB_BY_RESID_SCR,
    PD = 1/(1+16^2^((SCORE_CALC - 580)/20)),
    PD_TESLA = 1/(1+16^2^((SCORE_CALC - 20 - 580)/20)),
    PD_SME = 1/(1+16^2^((NOTE_SCORE - 580)/20)),
    PD_TESLA_SME = 1/(1+16^2^((NOTE_SCORE - 20 - 580)/20)))%>%
  mutate(DIFF_TIME = case_when( PAYMENT DATE == is.na(PAYMENT DATE) ~ diffTime(DUE DATE, Sys.Date(), units = "days"),
    TRUE ~ diffTime(DUE DATE, PAYMENT DATE, units = "days")),
    FPD_PAYMENT_FLAG = case_when(is.na(PAYMENT DATE) ~ "NO", TRUE ~ "YES")) %>%
  mutate(FPD_FLAG = case_when(DIFF_TIME < -30 ~ 1, TRUE ~ 0),
    FPD_FLAG_15 = case_when(DIFF_TIME < -15 ~ 1, TRUE ~ 0))
#####
```

Figure 18 – R : Nettoyage et transformation des données

6.3.3 Qualité des données

Il faut s'assurer que les données utilisées pour l'analyse sont **correctes, représentatives et fiables**.

Les premiers tests à effectuer portés sur **l'alignement temporel** pour veiller à ce que les données du risque par cohorte soient correctes. J'ai extrait mes données des +30@MOB6, +30@MOB3 et +30@MOB12 sur plusieurs mois de reporting et j'ai calculé leurs écarts par rapport aux données déjà existantes chez mon tuteur. J'ai réalisé ces vérifications sur Excel. Tant que je n'avais pas un écart nul, il fallait enquêter sur la source de l'erreur (souvent au niveau des jointures entre les tables sur SQL).

6.3.4 Modèle et Mesures

La troisième étape dans la production d'un rapport Power BI est la modélisation. Il s'agit de créer un modèle de données structuré et interconnecté à partir de différentes sources de données. Pour créer des liens entre elles, on crée des **relations de différentes cardinalités entre les tables qu'on a importé, en utilisant des clés primaires**.

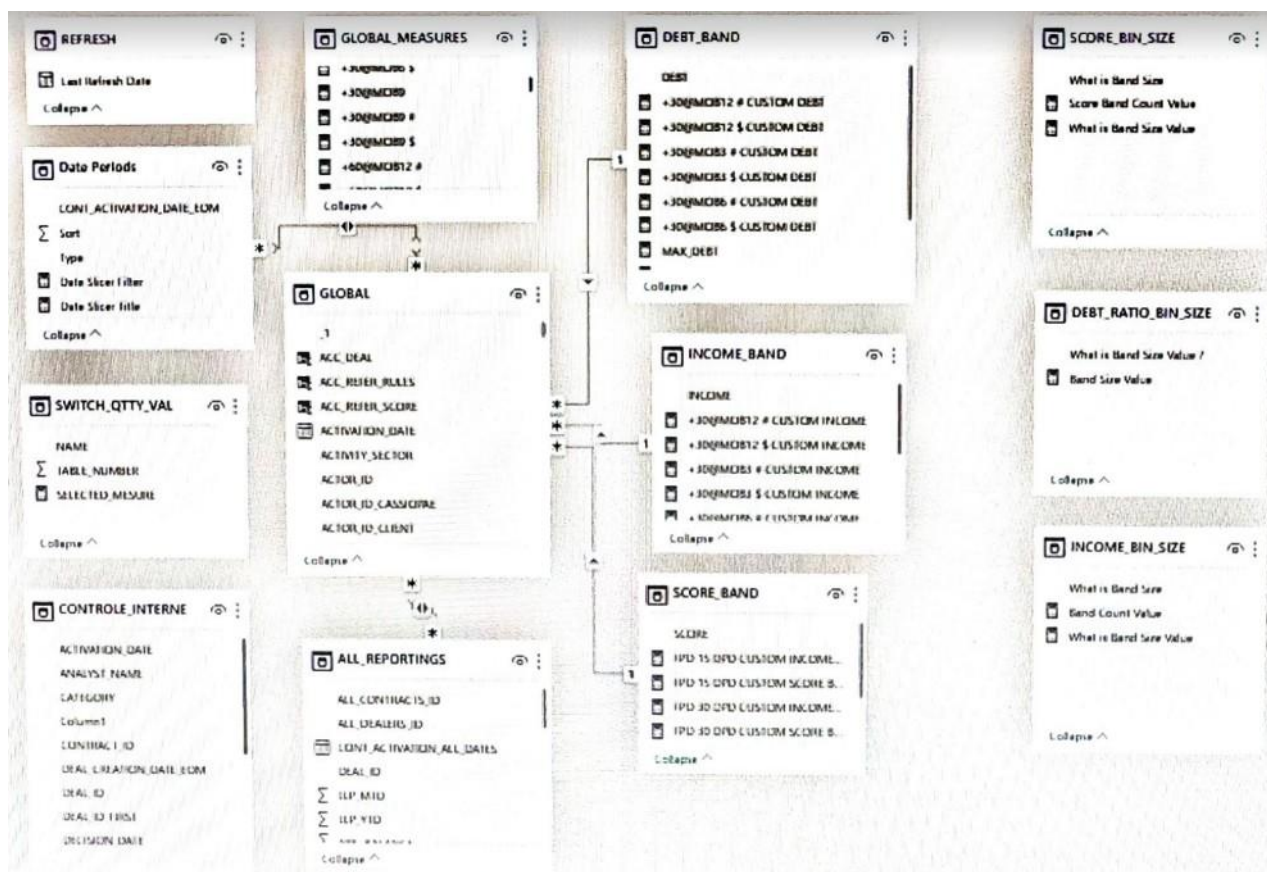


Figure 19 – Projet Power BI : Modélisation

Voici comment on était codé ces mesures sur Power BI en langage DAX.

MIN_SCORE	<code>MINX(ALL('SCORE_BAND'), 'SCORE_BAND'[SCORE])</code>
MAX_SCORE	<code>MAXX(ALL('SCORE_BAND'), 'SCORE_BAND'[SCORE])</code>
NBR CONTRACTS CUSTOM SCORE BIN	<pre> var _minScore = [MIN_SCORE] var _maxScore = [MAX_SCORE] var _bandSize = SCORE_BIN_SIZE / What is Band Size Value var _ScoreBandTable = GENERATESERIES(_minScore, _maxScore, _bandSize) var _currScore = SELECTEDVALUE('SCORE_BAND'[SCORE]) var _bandHead = MAXX(FILTER(_ScoreBandTable, [Value] <= _currScore), [Value]) var _bandTail = _bandHead + _bandSize return if(_currScore < _bandHead, CALCULATE(SUM('GLOBAL'[CONTRACT_QUANTITY]), 'SCORE_BAND'[SCORE] = _bandHead && 'SCORE_BAND'[SCORE] < _bandTail))</pre>
FPD IS DPD CUSTOM SCORE BIN \$	<pre> var _minScore = [MIN_SCORE] var _maxScore = [MAX_SCORE] var _bandSize = SCORE_BIN_SIZE / What is Band Size Value var _ScoreBandTable = GENERATESERIES(_minScore, _maxScore, _bandSize) var _currScore = SELECTEDVALUE('SCORE_BAND'[SCORE]) var _bandHead = MAXX(FILTER(_ScoreBandTable, [Value] <= _currScore), [Value]) var _bandTail = _bandHead + _bandSize return if(_currScore < _bandHead, CALCULATE(SUM('GLOBAL'[LOANAMOUNT]), 'GLOBAL'[FPD_FLAG] = 1, 'SCORE_BAND'[SCORE] = _bandHead && 'SCORE_BAND'[SCORE] < _bandTail))</pre>
FPD 30 DPD CUSTOM SCORE BIN \$	<pre> var _minScore = [MIN_SCORE] var _maxScore = [MAX_SCORE] var _bandSize = SCORE_BIN_SIZE / What is Band Size Value var _ScoreBandTable = GENERATESERIES(_minScore, _maxScore, _bandSize) var _currScore = SELECTEDVALUE('SCORE_BAND'[SCORE]) var _bandHead = MAXX(FILTER(_ScoreBandTable, [Value] <= _currScore), [Value]) var _bandTail = _bandHead + _bandSize return if(_currScore < _bandHead, CALCULATE(SUM('GLOBAL'[LOANAMOUNT]), 'GLOBAL'[FPD_FLAG] = 1, 'SCORE_BAND'[SCORE] = _bandHead && 'SCORE_BAND'[SCORE] < _bandTail))</pre>

Figure 20 – Projet Power BI : Regroupement dynamique

Malheureusement, je ne vais pas pouvoir montrer le résultat de cet outil. Cependant, je peux dire que j'ai réussi à atteindre la plupart des objectifs que nous nous sommes fixés, en termes de complétude et d'exactitude des données en backend.

7 Conclusion

Au cours de cette alternance chez Santander Consumer Banque, j'ai eu l'occasion de me plonger dans la gestion du risque de crédit et d'explorer en profondeur les processus décisionnels.

Les missions menées ont permis de participer activement à l'évolution des stratégies de gestion des risques, notamment par le biais de l'optimisation des modèles de scoring et l'intégration d'outils d'automatisation pour la surveillance des risques. Ces contributions ont été essentielles pour renforcer l'efficacité du système décisionnel de la banque, tout en répondant aux exigences de conformité et aux dynamiques du marché.

Les analyses des changements dans les Stratégies de Décision et de Scoring (SDS) ont mis en évidence la nécessité d'adapter constamment les outils et les méthodologies aux réalités économiques et aux régulations en vigueur. Les projets spécifiques menés, tels que l'amélioration des tableaux de bord Power BI et la gestion des risques de fraude, ont démontré l'importance de l'innovation et de l'agilité dans la gestion proactive des risques financiers.

Ce que j'ai le plus aimé, c'est la sensation d'être au carrefour entre les mathématiques appliquées, la finance et la technologie. J'ai pu fusionner des compétences techniques telles que la manipulation de données, la programmation et la visualisation avec des concepts financiers concrets comme les indicateurs de risque et les stratégies d'octroi. Cette conjonction d'aspects a renforcé mon sentiment de contribuer directement à la réussite de la banque.

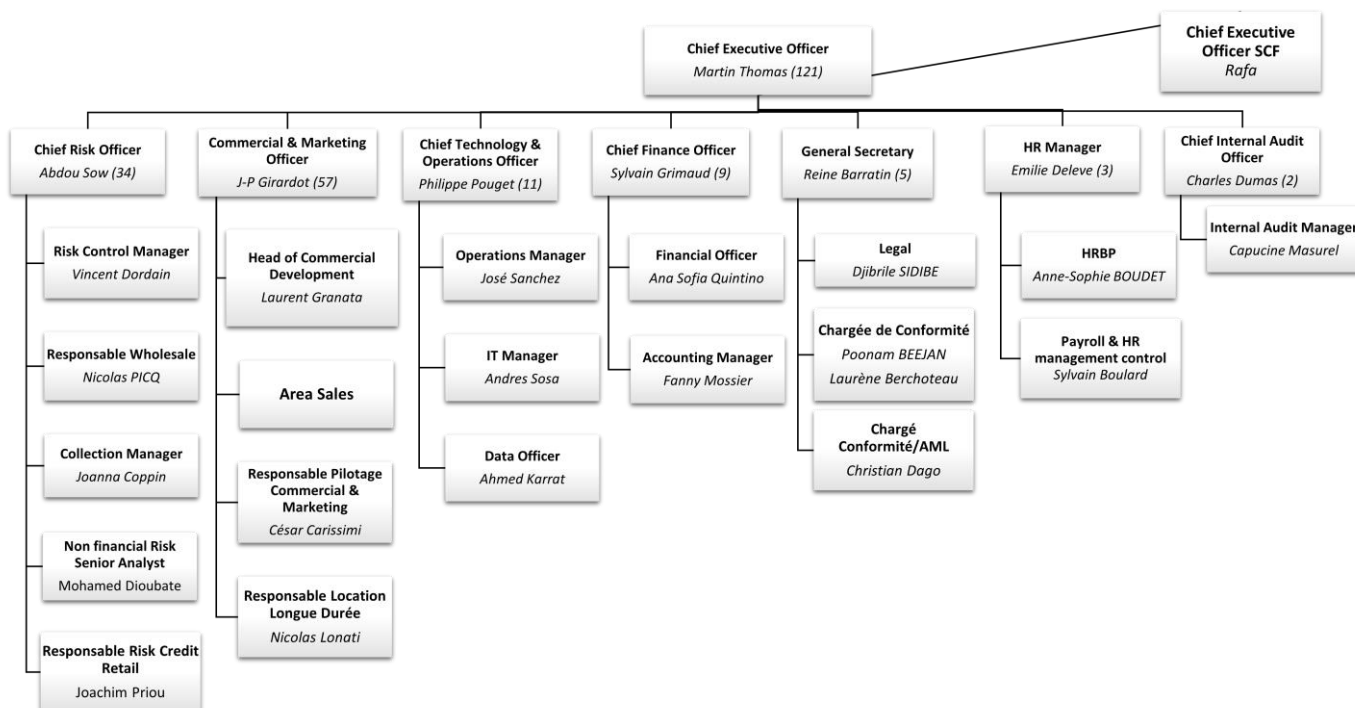
En conclusion, mon stage chez Santander Consumer Banque a été une expérience enrichissante. Il m'a permis de lier la théorie à la pratique, de développer de nouvelles compétences techniques et d'acquérir une perspective plus large sur le domaine de la finance et du risque. Je suis reconnaissant d'avoir eu l'opportunité de contribuer au sein d'une équipe aussi dynamique et motivée, et je suis convaincu que les connaissances et les compétences acquises durant ce stage continueront à me servir dans ma carrière future.

8 Bibliographie

- **Informations Internes - Santander Consumer Banque**
- **Site officiel de Santander Consumer Banque**
<https://www.santanderconsumer.fr/>
- **Comptes Sociaux de SCB 2021 - Papper.fr**
[SANTANDERCONSUMERBANQUE-Comptes sociaux 2021 16-05-2022. pdf](#)
- **What is risk management ? - IBM**
<https://www.ibm.com/topics/risk-management#:~:text=Risk%20analysis%20involves%20establishing%20the,according%20to%20prominence%20and%20consequence.>
- **Santander - Santander responds to the war in Ukraine - 07/03/2022**
<https://www.santander.com/en/press-room/press-releases/2022/03/santander-responses-to-the-war-in-ukraine>
- **Pourquoi faut-il se préoccuper de gérer les risques ? - ing.com** http://gpp.oiq.qc.ca/Start.htm#t=Pourquoi_faut-il_se_preoccuper_de_gerer_les_risques.htm
- **Le processus d'analyse et d'évaluation des risques - Ing. - Mai 2011**
http://gpp.oiq.qc.ca/Start.htm#t=Le_processus_d_analyse_et_d_evaluation_des_risques.htm

9 Annexes

Annexe I



- Organigramme de Santander Consumer Banque

Les étapes d'un développement rapide



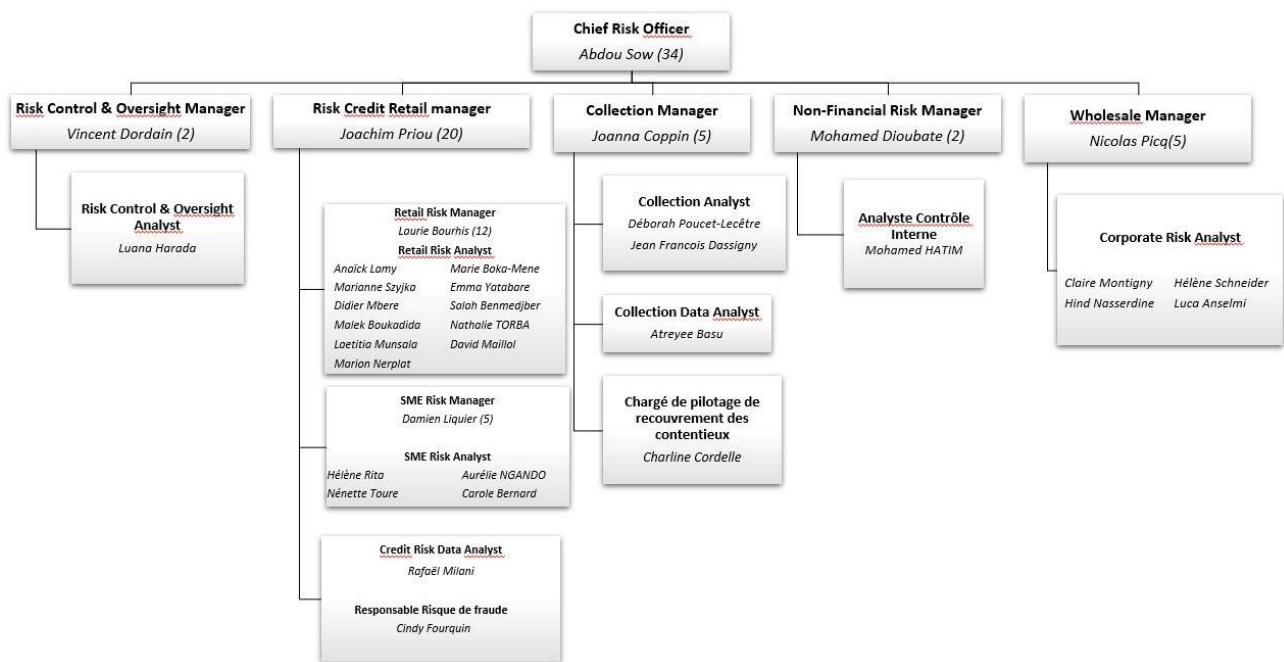
- Histoire de Santander Consumer Banque

Annexe II

<p>Vente à crédit (VAC)</p> <p>Définition : Crédit à la consommation consenti à une personne physique et destiné au financement de l'acquisition d'un bien mobilier, dont le remboursement se fait par versements périodiques.</p> <p>Client :</p> <ul style="list-style-type: none"> • Particulier • Entreprise (SME) <p>Propriétaire du véhicule : La personne physique (VAC particulier) ou morale (VAC SME) est juridiquement propriétaire de l'actif.</p>	<p>Location avec option d'achat (LOAC)</p> <p>Définition : La location avec option d'achat (LOA), encore appelée leasing, location avec promesse de vente ou crédit-bail, ne doit pas être confondue avec la location longue durée (LLD). En effet, les deux formules permettent de disposer d'un bien (souvent une voiture) sans l'acheter, en payant des loyers mensuels. Mais la LOA est un contrat de crédit qui offre la possibilité d'acheter le bien à la fin de la location, alors que la LLD est une simple location avec obligation de restitution.</p> <p>Client :</p> <ul style="list-style-type: none"> • Particulier <p>Propriétaire du véhicule : SCB est le propriétaire du véhicule. En cas d'exercice de l'option d'achat, la personne physique devient alors le propriétaire de l'actif.</p>	<p>Location avec option d'achat Ballon (LOAB)</p> <p>Définition : Même définition que LOAC. A la fin du contrat, le client est en droit d'exercer une option lui permettant de devenir propriétaire du véhicule. Si l'option n'est pas exercée, c'est au concessionnaire de reprendre le véhicule.</p> <p>Client :</p> <ul style="list-style-type: none"> • Particulier <p>Propriétaire du véhicule : SCB est le propriétaire du véhicule. La personne physique peut décider à l'issue du contrat d'exercer son option d'achat, et devient alors propriétaire de l'actif. Sinon le ballon est réglé par le concessionnaire qui devient propriétaire de l'actif.</p>
<p>Crédit Bail (CBAILC)</p> <p>Définition : SCB (le crédit-bailleur) met un bien (véhicules) à la disposition d'une entreprise pour une période déterminée, contre paiement d'une redevance périodique et d'un premier loyer souvent majoré.</p> <p>Au terme du contrat, l'entreprise bénéficiaire a généralement le choix entre plusieurs options : soit restituer le bien, soit l'acquiescer pour un montant défini lors de la conclusion du contrat, soit, plus rarement, le renouveler à des conditions le plus souvent moins coûteuses.</p> <p>Client :</p> <ul style="list-style-type: none"> • Entreprise (SME) <p>Propriétaire du véhicule : SCB est le propriétaire du véhicule.</p>	<p>Crédit Bail Ballon (CBAILB)</p> <p>Définition : Même définition que CBAILC. A la fin du contrat, le client est en droit de racheter le ballon lui permettant de devenir propriétaire du véhicule. Si l'option n'est pas exercée, c'est au concessionnaire de reprendre le véhicule.</p> <p>Client :</p> <ul style="list-style-type: none"> • Entreprise (SME) <p>Propriétaire du véhicule : SCB est le propriétaire du véhicule. La personne morale peut décider à l'issue du contrat de racheter le ballon défini à la signature, et devient alors propriétaire de l'actif. Sinon le ballon est réglé par le concessionnaire qui devient propriétaire de l'actif.</p>	<p>Crédit Stock</p> <p>Définition : Crédit de trésorerie confirmé ou non utilisable par tirages à CT, durée et taux fixés au départ de chaque tirage.</p> <p>Client :</p> <ul style="list-style-type: none"> • Concessionnaire (Wholesale) <p>Propriétaire du véhicule : Le concessionnaire est propriétaire des véhicules.</p>

- Produits de financement chez Santander Consumer Banque

Annexe III



- Organigramme de la Direction des Risques et du Recouvrement chez Santander ConsumerBanque

Annexe IV

BASES	TABLES	OVERVIEW	DESCRIPTION
DEVRIK	RAT_DEAL_EXPERIAN_HISTORY RDM_DIMENSIONS_DET RDM_PORT_EVOLUTION_DET	Données Score & Décision Infos Deals + Décisions Infos Contrats + Risk kpi	1 ligne par deal par appel de Score 1 ligne par deal 1 ligne par contrat par mois
	DATA_ACTOR_DIMENSIONS DATA_DEAL_DIMENSIONS DATA_CONT_PAYMENT_HISTORY DATA_CONT_WRTIEOFF_HISTORY DATA_ASSET_DIMENSIONS	Données Dealer Infos Deal + décisions Infos paiements Infos véhicule	1 ligne par dealer 1 ligne par dealer 1 ligne par contrat par mois 1 ligne par deal
	DATA_ARGUS_HISTORY MV_ND_RV_RISK_REPORT_ALL	N° des contrats de chez Argus Forecast Residual Value	1 ligne par contrat par mois 1 ligne par contrat par mois
DATARISK	THM_OUT_DEAL_JALON	Durée de chaque étape du process d'admission	1 ligne par étape par deal

- Descriptif des tables de données pour l'analyse de risque de crédit

Annexe V

+30DPD@MOB6 : Pourcentage de contrats ayant atteint 30 jours de retard de paiement au cours des 6 premiers mois de leur durée de vie.

$$\frac{\text{Nombre de contrats avec 30 DPD au MOB6}}{\text{Nombre total de contrats au MOB6}} \times 100$$

+30DPD@MOB12 : Pourcentage de contrats ayant atteint 30 jours de retard de paiement au cours des 12 premiers mois de leur durée de vie.

$$\frac{\text{Nombre de contrats avec 30 DPD au MOB12}}{\text{Nombre total de contrats au MOB12}} \times 100$$

NPL (Non-Performing Loans) : Pourcentage de prêts en défaut de paiement, c'est-à-dire les prêts sur lesquels les paiements ne sont pas effectués conformément aux termes du contrat.

$$\frac{\text{Montant total des prêts en défaut}}{\text{Montant total des prêts en portefeuille}} \times 100$$

FPD (First Payment Default) 15DPD : Pourcentage de contrats ayant atteint 15 jours de retard de paiement dès leur premier paiement.

$$\frac{\text{Nombre de contrats avec 15 DPD dès le premier paiement}}{\text{Nombre total de contrats dès le premier paiement}} \times 100$$

FPD30DPD : Pourcentage de contrats ayant atteint 30 jours de retard de paiement dès leur premier paiement.

$$\frac{\text{Nombre de contrats avec 30 DPD dès le premier paiement}}{\text{Nombre total de contrats dès le premier paiement}} \times 100$$

LLP (Loan Loss Provision) : Montant provisionné par la banque pour faire face aux pertes potentielles sur les prêts en défaut.

VMG (Volume of Managed Gross) : Volume brut des prêts sous gestion, utilisé pour calculer la base de prêt pour divers indicateurs de risque.

Annexe VI

```

#TEST REGRESSION DETAILED
orange_to_b_red <- before_db_d%>%filter(Output.Score >= 550 , Output.Score <= 553)%>%
  select(Input.Application.Number,INDEX, SEG, Output.Score, Output.Underwriting.Level, Output.Score.Decision,output.Final.Decision)

orange_to_b_green <- before_db_d%>%filter(Output.Score >= 561, Output.Score <=570)%>%
  select(Input.Application.Number,INDEX, SEG, Output.Score,output.Underwriting.Level, Output.Score.Decision,output.Final.Decision)

reg_red <- orange_to_b_red%>%inner_join(after_db_d%>%select(Input.Application.Number, INDEX,SEG, Output.Score,output.Underwriting.Level,
  Output.Score.Decision,output.Final.Decision),
  by=c("Input.Application.Number", "INDEX"))

reg_green <- orange_to_b_green%>%inner_join(after_db_d%>%select(Input.Application.Number,INDEX, SEG, Output.Score,output.Underwriting.Level,
  Output.Score.Decision,output.Final.Decision)
  ,by=c("Input.Application.Number", "INDEX"))

#orange to be red : only certain variables should change
var_3 <- before_db_d%>%filter(Output.Score >= 550 , Output.Score <= 553)%>%inner_join(after_db_d,
  by=c("INDEX", "Input.Application.Number"),
  suffix = c('_before', '_after'))

reg_var_3 <- difffdf(var_3[, endswith(colnames(var_3), "_before")]%>%
  rename_with(~str_remove(., '_before')),
  var_3[, endswith(colnames(var_3), "_after")]%>%
  rename_with(~str_remove(., '_after'))))

#orange to be green: only certain variables should change
var_4 <- before_db_d%>%filter(Output.Score >= 561, output.score <=570)%>%inner_join(after_db_d,
  by=c("INDEX", "Input.Application.Number"),
  suffix = c('_before', '_after'))

reg_var_4 <- difffdf(var_4[, endswith(colnames(var_4), "_before")]%>%
  rename_with(~str_remove(., '_before')),
  var_4[, endswith(colnames(var_4), "_after")]%>%
  rename_with(~str_remove(., '_after'))))

list_check <- list("NON_REG_ORANGE" = check_1, "NON_REG_RED_GREEN" = non_reg, "REG_2B_RED"= reg_red, "REG_2B_GREEN" = reg_green)
require(openxlsx)

write.xlsx(list_check , "/csv/tasks/impact_score/checks_reg_non_reg/checks_impact_auto_sup_6k.xlsx")

```

- Score Particulier : Tests de régression sur R

REGIMENT	CHECK	DESCRIPTION	NON REGRESSION	OUTPUT	EXPLICATIONS	CHECK	DESCRIPTION	REGRESSION	OUTPUT
RISK SCORE AUTODECISION	1	Déclat HORS TESLA ET NON EXISTING CUSTOMERS oranges non concernés inchangés	NON REG	1	Déclat passant de Orange à Rouge UNIQUEMENT	1	Déclat passant de Orange à Rouge UNIQUEMENT	REG	1
	2	Déclat HORS TESLA et NON EXISTING CUSTOMERS rouges/verts non concernés inchangés	NON REG	2	Déclat passant de Orange à Vert UNIQUEMENT	2	Déclat passant de Orange à Vert UNIQUEMENT	REG	2

	CHECKS	NON REGRESSION				REGRESSION			
		#ANALYZED	#KO	% KO		#ANALYZED	#KO	% KO	
SCORE AUTODECISION >= 6K	1		2388	0	0,00%	1278	0	0,00%	
	2		27752	0	0,00%	3405	0	0,00%	
SCORE AUTODECISION <6K moto	1		6760	0	0,00%	227	0	0,00%	
	2	NA				419	0	0,00%	
TESLA SCORE FRAUDE	1		51367	0	0,00%	5675	0	0,00%	
	2	NA					NA		
TESLA SCORE AUTODECISION	1		2834	0	0,00%	130	0	0,00%	
	2		89	0	0,00%	135	0	0,00%	

Check	Non regression	Regression
1	Oranges non concernées restent oranges	Oranges qui devrait devenir ROUGE le deviennent
2	Verts/rouges non concernées restent verts/rouges	Oranges qui devrait devenir VERTS le deviennent
3	Existing Customers ne sont pas en SUSP FRAUDE	

- Aperçu d'un Data-Testing Log