

TheLook Fintech

Resumo Técnico do Projeto

Este projeto teve como objetivo principal apoiar o departamento financeiro da TheLook Fintech com análises baseadas em dados, desde a ingestão e tratamento das informações até a geração de relatórios e dashboards estratégicos. O trabalho foi conduzido com foco em entregar respostas a perguntas-chave do negócio, promovendo uma cultura de dados orientada à tomada de decisões.

Principais atividades realizadas:

- Coleta, integração e análise de **dados brutos de múltiplas fontes** (internas e externas).
- Processamento, limpeza e modelagem de dados no ambiente de **nuvem com Google BigQuery**.
- Construção de **consultas SQL otimizadas** para agregações, filtros, joins e criação de tabelas analíticas.
- Desaninhamento e manipulação de colunas do tipo **RECORD** (structs).
- Desenvolvimento de relatórios e **visualizações interativas com o Looker (Looker Studio/Looker Enterprise)**.
- Criação de **dashboards de autoatendimento** para acompanhamento do portfólio de empréstimos.
- Implementação de funcionalidades como **formatação condicional, cruzamento de filtros e atualizações automáticas** nos dashboards.
- Compartilhamento de resultados com stakeholders e sugestões de evolução para os processos analíticos da fintech.

Linguagens e tecnologias utilizadas:

- **SQL (BigQuery SQL)** – para exploração, transformação e modelagem dos dados.
- **Google BigQuery** – como engine de processamento e armazenamento em nuvem.

- **Google Cloud Storage (GCS)** – para ingestão de arquivos externos em formato **.CSV**.
- **Looker (Looker Studio/Enterprise)** – para construção dos dashboards interativos.
- **Planilhas Google** – integração leve para compartilhamento de dados tabulares.
- **Google Cloud Console / GCP Tools** – ambiente geral de operação e controle de permissões.

Estrutura e entregas do projeto:

- 5 fases organizadas por escopo analítico:
 1. Exploração dos dados
 2. Integração externa e modelagem relacional
 3. Análise de propósitos e risco
 4. Relatórios financeiros agregados
 5. Criação e entrega de dashboard interativo
- 7 perguntas-chave de negócio respondidas
- Propostas estruturadas de **melhorias analíticas**, **modelos de machine learning futuros** e **expansão da inteligência de dados** para outras áreas da empresa

Cenário

Atuação como analista de dados em nuvem na TheLook Fintech, uma startup em expansão no setor de tecnologia financeira que oferece empréstimos a proprietários de lojas on-line independentes, com o objetivo de facilitar o acesso a crédito para pequenos empreendedores que desejam investir em estoque e expandir seus negócios.

Com o mercado-alvo já bem definido, a empresa vive um momento de crescimento acelerado e demanda por decisões cada vez mais estratégicas e orientadas por dados. Nesse contexto, fui

responsável por apoiar o departamento financeiro no uso eficiente de dados para acompanhar o desempenho da operação e sustentar a tomada de decisões críticas.

A atribuição principal inicial foi desenvolver e implementar um plano de análise com foco em três questões de negócio identificadas em reunião com a liderança do setor:

- **Como monitorar melhor o fluxo de caixa**, garantindo que o volume de empréstimos liberados mensalmente não ultrapasse os recursos que entram no caixa?
- **Como identificar os principais motivos que levam os clientes a tomar empréstimos conosco?**
- **Como rastrear os locais onde os mutuários estavam no momento da contratação dos empréstimos?**

Durante as conversas com o líder financeiro, Tiago, foram definidos os principais indicadores a serem considerados na análise. O fluxo de caixa — representando entradas e saídas de recursos — precisa ser constantemente monitorado para assegurar a sustentabilidade da operação. O **propósito do empréstimo** também é uma métrica crítica, pois existe uma forte correlação entre a razão do empréstimo e a probabilidade de quitação. Já a **localização geográfica dos mutuários** se mostrou essencial para avaliar a exposição ao risco: uma alta concentração de empréstimos em uma única região pode aumentar o risco de inadimplência coletiva, ao passo que uma distribuição mais equilibrada entre diferentes áreas reduz essa vulnerabilidade.

A análise foi, portanto, direcionada para essas três dimensões fundamentais: fluxo de caixa, propósito do empréstimo e localização dos mutuários, como base para a construção de relatórios e painéis capazes de oferecer uma visão estratégica sobre a carteira da empresa.

Fases do Projeto

Fase 1 - Análise inicial dos dados

Nesta etapa inicial, o foco foi me familiarizar com o ambiente de trabalho no Google BigQuery e explorar o conjunto de dados da TheLook Fintech. O objetivo foi entender a estrutura, o conteúdo e a qualidade dos dados disponíveis, a fim de preparar o terreno para as análises posteriores.

As atividades realizadas incluíram:

- **Abertura do ambiente de BigQuery e identificação do projeto** onde os dados estavam armazenados.
- **Localização do dataset `fintech`** na seção “Explorer” do BigQuery, que concentra todas as tabelas relevantes para a análise do negócio.
- **Exploração detalhada de cada tabela do conjunto de dados**, utilizando as abas:
 - **Esquema (Schema)**: para identificar as colunas disponíveis e seus respectivos tipos de dados.
 - **Detalhes (Details)**: para obter informações adicionais sobre cada tabela.
 - **Visualização (Preview)**: para observar uma amostra dos dados e verificar sua consistência.

Durante essa análise exploratória, identifiquei elementos essenciais para responder às perguntas de negócio definidas com o time financeiro. Alguns pontos-chave:

- A coluna `loan_amount` da tabela `loan` foi confirmada como a fonte principal para análise dos valores concedidos em empréstimos.
- A coluna `issue_d` da tabela `loan` foi identificada como a data de emissão dos empréstimos — informação essencial para monitorar o fluxo de saída de capital.
- A coluna `int_rate` da tabela `loan` fornece a taxa de juros aplicada a cada empréstimo.
- Outras colunas e tabelas relevantes, como `customer`, `sales` e `products`, também foram inspecionadas para futuras correlações.

Essa etapa foi fundamental para mapear os dados disponíveis e estabelecer uma base sólida para as próximas fases do projeto, garantindo que as decisões analíticas estejam alinhadas à estrutura real do banco de dados da empresa.

Fase 2 – Integração de dados externos e modelagem com SQL

Nesta etapa, o foco foi integrar dados externos ao conjunto principal de informações da empresa com o objetivo de **rastrear os mutuários por localização geográfica**, fator essencial para monitoramento de riscos de inadimplência concentrada. Essa atividade se conecta diretamente à pergunta de negócio: **“Quais locais (estados ou regiões) concentram os empréstimos concedidos?”**

Importação de dados externos via CSV

Foi realizado o carregamento de um arquivo `.csv` hospedado no Google Cloud Storage contendo o mapeamento de **estados norte-americanos para regiões e sub-regiões**. Esse arquivo foi importado para o BigQuery utilizando o comando abaixo, criando a tabela `state_region` no dataset `fintech`:

```
LOAD DATA OVERWRITE fintech.state_region

(
  state STRING,

  subregion STRING,

  region STRING
)

FROM FILES (
  format = 'CSV',

  uris =
  ['gs://sureskills-lab-dev/future-workforce/da-capstone/temp_35_us/stat
e_region_mapping/state_region_*.csv']
);
```

Junção de tabelas com INNER JOIN

Para consolidar as informações de localização dos mutuários, foi realizada uma junção entre a tabela `loan` e a recém-importada `state_region`, utilizando o campo `state` como chave de ligação:

```
SELECT

  lo.loan_id,

  lo.loan_amount,

  sr.region
```

```
FROM fintech.loan lo

INNER JOIN fintech.state_region sr

ON lo.state = sr.state;
```

Figura 1 – Dado externo CSV importado

[Visualizar imagem no GitHub](#)

Essa junção permitiu associar cada empréstimo a sua respectiva região, enriquecendo a base para análises geográficas e estratégicas.

Criação de tabela com CTAS (CREATE TABLE AS SELECT)

Como resposta estruturada à demanda de rastreamento dos mutuários por região, foi criada uma tabela permanente utilizando a instrução CTAS:

```
CREATE OR REPLACE TABLE fintech.loan_with_region AS

SELECT

    lo.loan_id,

    lo.loan_amount,

    sr.region

FROM fintech.loan lo

INNER JOIN fintech.state_region sr

ON lo.state = sr.state;
```

Figura 2 – Query criação tabela loan_with_region

[Visualizar imagem no GitHub](#)

Figura 3 – Preview da tabela loan_with_region

[Visualizar imagem no GitHub](#)

Adicionais sugeridos

Além do atendimento direto à pergunta de negócio, sugeri o desenvolvimento de **agregações por região**, aplicando `COUNT(*)` para identificar regiões com maior volume de empréstimos. Essas contagens podem ser ordenadas para gerar um **ranking de regiões críticas**, representando um passo preliminar para relatórios de risco geográfico e priorização de estratégias de mitigação.

Fase 3 – Relatório de propósito por meio dos dados aninhados

Esta fase responde diretamente à pergunta de negócio: “**Quais são os principais motivos que levam os clientes a pegarem empréstimos?**”, destacando-se como uma análise qualitativa da carteira.

Desaninhamento de dados

Os dados sobre o propósito de cada empréstimo estavam armazenados em **colunas aninhadas** dentro de um campo do tipo `RECORD` chamado `application`. Para extrair esses dados, foi necessário utilizar notação de ponto para acessar subcampos:

```
SELECT loan_id, application.purpose
FROM fintech.loan;
```

Essa extração possibilitou visualizar e analisar os motivos declarados pelos clientes no momento da solicitação dos empréstimos, contribuindo para estudos de perfil de comportamento e risco.

Eliminação de duplicatas e criação de CTAS para relatórios

Foi observado que muitos propósitos se repetiam na base. Para construir uma base limpa e utilizável em relatórios, foi aplicada a eliminação de duplicatas com `DISTINCT`, e em seguida, criado um CTAS contendo somente os propósitos únicos:

```
CREATE OR REPLACE TABLE fintech.loan_purposes AS
SELECT DISTINCT application.purpose AS purpose
FROM fintech.loan;
```

Essa tabela serve como base para painéis qualitativos ou categorização de riscos por finalidade.

Figura 4 – Tabela **loan_purposes** criada após eliminação de duplicatas

[Visualizar imagem no GitHub](#)

Adicionais sugeridos

- Criação de **agregações com COUNT(*) GROUP BY application.purpose**, para identificar os propósitos mais frequentes entre os mutuários.

Exemplo:

```
SELECT application.purpose, COUNT(*) AS total
FROM fintech.loan
GROUP BY application.purpose
ORDER BY total DESC;
```

Figura 5 – Query adicional com contagem de empréstimos por propósito

[Visualizar imagem no GitHub](#)

- Inclusão de uma **coluna adicional com classificação de risco por propósito**, categorizando os motivos entre "preocupantes para o negócio" e "tranquilos para o negócio", com base em critérios definidos pela inteligência de negócios da empresa. Essa classificação poderá ser utilizada futuramente para gerar **relatórios de propensão à inadimplência por motivo declarado**, auxiliando nas estratégias de concessão e precificação de crédito.

Fase 4 – Geração de relatórios agregados com SQL para insights financeiros

Nesta fase, concentrei os esforços em criar relatórios consolidados com base em dados históricos de empréstimos, com foco em oferecer **insights objetivos para o monitoramento do fluxo de caixa da empresa**. A etapa responde diretamente à pergunta de negócio: “**Como**

monitorar melhor o fluxo de caixa para garantir que o volume de empréstimos liberados a cada mês não supere o montante que entra no caixa?”

Consolidação do valor total de empréstimos por ano

A primeira análise consistiu em calcular o valor total de empréstimos emitidos por ano. Isso permite à equipe financeira identificar tendências no volume de capital liberado e entender como esse comportamento evolui ao longo do tempo. A consulta utilizada foi:

```
SELECT  
  
    issue_year,  
  
    SUM(loan_amount) AS total_amount  
  
FROM fintech.loan  
  
GROUP BY issue_year  
  
ORDER BY issue_year;
```

Esse relatório fornece uma visão clara da variação anual nos empréstimos concedidos, sendo uma métrica direta da saída de caixa da operação.

Figura 6 – Query SQL e resultado com soma dos valores de empréstimos por ano

[Visualizar imagem no GitHub](#)

Contagem de empréstimos por ano com CTAS

Para complementar a análise de volume financeiro, foi criado um relatório com o total de empréstimos emitidos por ano, utilizando a função `COUNT()` sobre o `loan_id`:

```
CREATE OR REPLACE TABLE fintech.loan_count_by_year AS  
  
SELECT  
  
    issue_year,  
  
    COUNT(loan_id) AS total_loans
```

```
FROM fintech.loan

GROUP BY issue_year

ORDER BY issue_year;
```

Essa tabela foi incorporada ao pipeline de dados para permitir a comparação entre o número de operações realizadas e o montante financeiro movimentado, fornecendo indicadores como ticket médio por empréstimo e sazonalidade operacional.

Esses relatórios agregados representam um primeiro passo importante na construção de inteligência financeira baseada em dados históricos.

Figura 7 – Query de criação da tabela **loan_count_by_year**

[Visualizar imagem no GitHub](#)

Figura 8 – Visualização da tabela **loan_count_by_year** criada

[Visualizar imagem no GitHub](#)

Além disso, como forma de **aprofundar a análise financeira**, sugeri a criação de novos relatórios complementares com partições mais refinadas de tempo, como:

- **Valor total de empréstimos por mês**, para permitir acompanhamento mais granular e detectar variações mensais de comportamento.
- **Agrupamentos por trimestre e semestre**, úteis para relatórios gerenciais e alinhamento com ciclos orçamentários.
- **Ranking dos meses de maior volume liberado**, identificando sazonalidades e momentos de pico de risco financeiro, fornecendo à empresa insumos valiosos para planejamento de liquidez e política de crédito.

Essas iniciativas, integradas aos dashboards e rotinas de acompanhamento, agregam valor analítico ao setor financeiro e permitem decisões mais fundamentadas com base em comportamento histórico e tendências operacionais.

Atualização de Cenário

Com a evolução do trabalho analítico e os relatórios entregues nas fases anteriores, a equipe financeira da TheLook Fintech passou a ter acesso a dados mais estruturados e confiáveis para acompanhar o desempenho da operação. A partir desses resultados, novas necessidades surgiram — e com elas, uma nova demanda liderada por Trevor, responsável pelo departamento de gestão financeira da empresa.

Trevor solicitou o desenvolvimento de um **dashboard interativo**, com o objetivo de oferecer **visibilidade rápida e prática** sobre os dados de **integridade dos empréstimos**, um tema central nas discussões da equipe executiva. A ideia era transformar as análises em visualizações acessíveis, integradas e acionáveis.

Durante o alinhamento com Trevor, quatro **perguntas de negócio prioritárias** foram identificadas:

1. **Qual é o valor total pendente de todos os empréstimos?**
2. **Qual é a porcentagem de empréstimos pendentes em cada categoria de status?**
3. **Quais estados têm a maior quantidade de empréstimos pendentes?**
4. **Quais clientes já quitaram a própria casa e têm empréstimos vigentes?**

Essas questões estão diretamente ligadas a **métricas críticas para avaliação de risco** e compliance regulatório. O **saldo total de empréstimos pendentes**, por exemplo, reflete a exposição financeira da empresa; o **status do empréstimo** permite identificar atrasos, inadimplência e padrões de pagamento; a **distribuição geográfica** dos contratos ajuda a mitigar riscos concentrados; e a **análise de renda vs. propriedade de imóvel** contribui para decisões mais justas e equilibradas nas ofertas de crédito.

Com base nesses direcionamentos, iniciei a construção do dashboard, que passou a representar uma nova fase do projeto — consolidando o papel da análise de dados como suporte direto à estratégia da empresa.

Fase 5 – Criação de Dashboard Estratégico para Monitoramento da Integridade dos Empréstimos

Com base nas novas demandas apresentadas pelo departamento financeiro da TheLook Fintech, iniciei a construção de um dashboard interativo na plataforma Looker, com o objetivo de consolidar visualmente os principais indicadores relacionados à integridade dos empréstimos. A criação do painel partiu da pasta pessoal do projeto (Developer Student), onde

foi criado o dashboard nomeado "**Insights sobre empréstimos**", preparado para receber as visualizações solicitadas.

A seguir, detalho as visualizações criadas e suas respectivas lógicas analíticas.

Visualização 1 – Valor Total de Empréstimos Pendentes

Essa visualização responde à pergunta: "**Qual é o valor total pendente de todos os empréstimos?**" O foco foi apresentar, de forma direta, o montante financeiro total ainda não quitado pelos mutuários.

Figura 9 – Conferência da modelagem LookML

[Visualizar imagem no GitHub](#)

Utilizei a medida `Outstanding Loans Amount` do modelo `loan.view` e apliquei o tipo de visualização **Single Value**, ideal para destacar um indicador-chave. Para reforçar a interpretação, foi adicionada **formatação condicional**: o valor é destacado em vermelho caso ultrapasse **US\$ 3.000.000.000**, limite estabelecido pela equipe como parâmetro de exposição de risco.

Figura 10 – Formatação condicional para visualização 1

[Visualizar imagem no GitHub](#)

SQL subjacente à visualização:

```
SELECT
  COALESCE(SUM(CASE
    WHEN (loan.loan_status <> 'Fully Paid' OR loan.loan_status IS
NULL)
    THEN loan.loan_amount
    ELSE NULL
  END), 0) AS loan_outstanding_loans_amount
FROM `cloud-training-demos.fintech.loan` AS loan
LIMIT 500;
```

Figura 11 – Visualização 1 finalizada

[Visualizar imagem no GitHub](#)

Visualização 2 – Porcentagem de Empréstimos Pendentes por Status

Esta visualização responde à pergunta: **"Qual é a porcentagem de empréstimos pendentes em cada categoria de status?"** Ela fornece uma leitura da integridade geral da carteira, apontando proporções de empréstimos em situação de atraso, inadimplência, carência ou ativos.

Foi utilizada a dimensão **Loan Status** combinada com a medida **Count**, exibida em um **gráfico de pizza (Pie Chart)**, facilitando a compreensão da distribuição percentual entre os diferentes status de empréstimos.

Figura 12 – Visualização 2 (preview)

[Visualizar imagem no GitHub](#)

Visualização 3 – Contagem Total de Empréstimos Pendentes por Estado

Atendendo à demanda de visualizar a **concentração geográfica dos empréstimos pendentes**, esta visualização responde à pergunta: **"Quais estados têm a maior quantidade de empréstimos pendentes?"**

Foi utilizada a dimensão **State** com a medida **Count**. A visualização foi estruturada em **gráfico de barras horizontais**, limitada aos **10 estados com maior volume de empréstimos em aberto**, para permitir rápida identificação de concentrações de risco por região.

Figura 13 – Visualização 3 (preview)

[Visualizar imagem no GitHub](#)

Visualização 4 – Dez Principais Clientes por Maior Renda

Por fim, esta visualização responde à pergunta: "**Quais clientes já quitaram a própria casa e têm empréstimos vigentes?**", filtrando os 10 clientes com maior renda individual.

Foram utilizadas quatro dimensões:

- `customer_id`
- `annual_income`
- `state`
- `interest_rate`

A visualização foi organizada em **formato de tabela ordenável**, permitindo análise detalhada dos perfis de clientes de maior potencial financeiro, com filtro aplicado para exibir apenas aqueles que **já quitaram sua residência e possuem empréstimos ativos**.

Figura 14 – Quatro visualizações prontas no dashboard

[Visualizar imagem no GitHub](#)

Funcionalidades Adicionais, Otimização e Design do Dashboard

Além das visualizações principais, foram aplicadas melhorias funcionais e de usabilidade ao dashboard, visando melhor performance e experiência para o time:

- **Cruzamento de filtros (cross-filtering):** habilitado para permitir filtragem dinâmica entre blocos, de acordo com a seleção feita em qualquer visualização.
- **Atualizações automáticas por bloco:**
 - "Valor total dos empréstimos pendentes": atualizado **a cada hora**
 - "Porcentagem de empréstimos pendentes" e "Dez principais clientes por renda mais alta": atualizados **diariamente**

- **Design responsivo e organizado:** layout distribuído com agrupamento lógico de visualizações, títulos claros, paleta de cores funcional e navegação intuitiva.

Figura 15 – Configuração das atualizações automáticas por tempo

[Visualizar imagem no GitHub](#)

Compartilhamento com o Time

Finalizada a construção e validação do dashboard, gerei o link de acesso via Looker e compartilhei com Trevor e os principais stakeholders da área financeira. A disponibilização do painel permitiu que a equipe acompanhasse, de forma contínua e visual, os indicadores essenciais para a tomada de decisão estratégica da empresa.

Figura 16 – Dashboard final com visualizações e funcionalidades concluídas

[Visualizar imagem no GitHub](#)

Sugestões adicionais

Como propostas para evolução do dashboard e aprimoramento da inteligência operacional e estratégica, foram sugeridas as seguintes implementações futuras:

- **Criação de um sistema automatizado de alertas** com base na Visualização 1, para notificar o time sempre que o valor total de empréstimos pendentes ultrapassar uma determinada porcentagem do limite de US\$ 3.000.000.000 estabelecido pela equipe financeira.
- **Alerta de concentração de risco geográfico**, com foco em regiões que apresentem aumento significativo no volume de empréstimos, potencializando o risco de inadimplência coletiva.
- **Aplicação de inteligência de marketing integrada ao painel**, sugerindo ações direcionadas como ofertas personalizadas de melhores taxas ou novos empréstimos para clientes bem avaliados. Essa recomendação seria baseada em filtros selecionados pelo time financeiro, como renda, histórico de pagamento ou status de propriedade.

Essas funcionalidades têm o potencial de transformar o dashboard de uma ferramenta analítica passiva em um recurso proativo de **monitoramento, mitigação de risco e estímulo à retenção e expansão da base de clientes**.

Conclusão

Este projeto representou um ciclo completo de análise e entrega de valor baseado em dados, desde a exploração e modelagem até a construção de relatórios e visualizações estratégicas. Ao longo do processo, foram utilizadas ferramentas como **BigQuery** e **Looker**, combinadas com técnicas de **modelagem relacional**, **agregações SQL** e **análise exploratória**, para responder de forma objetiva às perguntas críticas do departamento financeiro da TheLook Fintech.

A seguir, destaco os **resultados alcançados**, organizados com base nas **perguntas de negócio iniciais** e os **impactos imediatos gerados para a empresa**:

1. Como monitorar melhor nosso fluxo de caixa?

- **Solução:** Criação de relatórios agregados com o valor total de empréstimos emitidos por ano e a contagem de empréstimos por período.
 - **Resultado para o negócio:** Visibilidade clara sobre a **saída de capital ao longo do tempo**, permitindo **previsões de liquidez** e **ajuste de políticas de concessão**.
 - **Métricas futuras:** Relatórios mensais, trimestrais e rankings de meses com maior liberação de crédito.
-

2. Como identificar os principais motivos que levam os clientes a pegar empréstimos conosco?

- **Solução:** Análise de dados aninhados e criação de tabela com os propósitos dos empréstimos.
 - **Resultado para o negócio:** Entendimento da **correlação entre o motivo do empréstimo e o risco de inadimplência**, auxiliando na **personalização de produtos financeiros**.
 - **Métricas futuras:** Agregações por volume de empréstimos por motivo e coluna de **classificação de risco** por tipo de propósito.
-

3. Como rastrear os locais onde os mutuários estavam quando contrataram os empréstimos?

- **Solução:** Integração de dados externos via CSV e associação geográfica via SQL.
 - **Resultado para o negócio:** Capacidade de identificar **regiões com alta concentração de empréstimos** e mitigar **riscos de inadimplência coletiva**.
 - **Métricas futuras:** Alertas automáticos para **concentração de risco por região** e filtros dinâmicos por área crítica.
-

4. Qual é o valor total pendente de todos os empréstimos?

- **Solução:** Visualização direta do saldo pendente total com formatação condicional.
 - **Resultado para o negócio:** Monitoramento contínuo da **exposição total da empresa ao risco de crédito**.
 - **Aditivo sugerido:** Criação de **alertas automáticos** quando valores ultrapassarem limites estabelecidos.
-

5. Qual é a porcentagem de empréstimos pendentes em cada categoria de status?

- **Solução:** Visualização percentual (gráfico de pizza) por categoria de status de empréstimos.
 - **Resultado para o negócio:** **Visão clara da saúde da carteira**, com identificação de crescimento de inadimplência e planejamento de ações corretivas.
 - **Métrica futura:** Acompanhamento histórico da evolução de cada status trimestralmente.
-

6. Quais estados têm a maior quantidade de empréstimos pendentes?

- **Solução:** Ranking visual dos 10 estados com maiores volumes pendentes.
 - **Resultado para o negócio:** Facilita **decisões de priorização geográfica** e estratégias de mitigação.
 - **Aditivo sugerido:** Gatilhos para **alertas regionais** e **ações específicas por território**.
-

7. Quais clientes já quitaram a própria casa e têm empréstimos vigentes?

- **Solução:** Tabela com os 10 principais clientes por renda e perfil patrimonial.
- **Resultado para o negócio:** Identificação de **clientes de baixo risco** para **novas oportunidades de crédito**, upsell e fidelização.
- **Aditivo sugerido:** Integração entre **marketing e crédito** para envio de propostas personalizadas.
-

Propostas para Evolução e Projetos Futuros

A partir dos aprendizados e resultados obtidos neste projeto, e com base na minha experiência como analista e cientista de dados, identifiquei oportunidades de evolução que podem ampliar significativamente o impacto da área de dados sobre as decisões da TheLook Fintech. As ideias abaixo são propostas realistas, escaláveis e alinhadas com as demandas de uma startup em crescimento, especialmente no setor financeiro.

1. Modelo de Credit Score com Machine Learning

Desenvolvimento de um modelo preditivo de **score de crédito** baseado em dados históricos de empréstimos, utilizando algoritmos como **LightGBM ou XGBoost**. O objetivo é melhorar a acurácia na avaliação de risco de crédito e personalizar limites e condições para cada cliente. O modelo pode ser integrado aos sistemas operacionais da fintech para automatizar e escalar decisões de concessão.

2. Integração de GenAI com Modelo de Crédito

Implementação de um **chatbot com GenAI**, acoplado ao modelo de score de crédito, para oferecer em tempo real a melhor taxa possível de empréstimo para o vendedor na ponta, com

base em seu perfil e contexto. A solução traria agilidade, personalização e transparência para o processo de contratação, além de otimizar margens operacionais.

3. Modelo de Churn para Risco de Inadimplência

Construção de um modelo supervisionado (Decision Tree Classifier, XGBoost ou similares) para **prever o risco de churn/inadimplência**, identificando perfis com maior probabilidade de não quitação. O modelo pode acionar ações preventivas, renegociação antecipada ou bloqueio de crédito, reduzindo perdas financeiras.

4. Geração de Leads Qualificados via Dados Públicos

Criação de dashboards e relatórios automatizados para mapeamento de **novos negócios** com base em ingestão de dados de fontes governamentais (ex.: abertura de empresas, alterações cadastrais e faturamento declarado). A ideia é gerar uma base ativa de **leads qualificados**, segmentados por porte, setor e localização, potencializando ações comerciais.

5. Cultura de dados e Self-Service Analytics

Implementação de um sistema de **self-service de dados** acessível a todas as áreas da empresa, com dashboards padronizados e treinamento interno da equipe. Essa medida visa descentralizar a análise operacional, promovendo autonomia dos times, reduzindo gargalos e reforçando a cultura de decisões orientadas por dados.

Encerramento

Este projeto não apenas atendeu às perguntas de negócio inicialmente propostas — envolvendo fluxo de caixa, rastreamento de mutuários, propósitos de empréstimo e integridade da carteira — como também estabeleceu bases sólidas para o crescimento da inteligência analítica dentro da TheLook Fintech.

Com as ferramentas, práticas e sugestões apresentadas, a empresa está preparada para evoluir de uma operação baseada em dados descritivos para uma abordagem cada vez mais preditiva e estratégica. A ciência de dados tem potencial de se tornar um dos principais diferenciais competitivos da fintech — e este projeto representa um ponto de partida concreto nesse caminho.

Estou à disposição para participar de novos ciclos de desenvolvimento, contribuindo com a expansão das soluções de dados e inteligência aplicada em todas as áreas da organização.