
RELATÓRIO DE ANÁLISE DE DADOS BANCÁRIOS PARA INCIDIUM

1 Introdução

Com a transformação digital, a Análise de Dados Bancários tornou-se imprescindível, permitindo identificar padrões, tendências e correlações em grandes volumes de dados transacionais. A inovação do Onboarding Digital está remodelando as práticas tradicionais de análise. A análise de dados revoluciona a operação dos bancos, possibilitando previsões mais precisas e decisões baseadas em dados. Isso inclui prever comportamentos de clientes, inadimplência, atividades fraudulentas e melhorar a satisfação do cliente. A previsão de riscos é um aspecto essencial, permitindo aos bancos gerenciar melhor os recursos e minimizar riscos relacionados a empréstimos e investimentos. Além disso, a análise de dados permite a personalização de produtos, oferecendo serviços sob medida com base no comportamento e histórico do cliente. Também contribui para o gerenciamento de conformidade, ajudando a identificar atividades suspeitas e garantir aderência a regulamentações. No dia a dia do banco, a análise de dados responde a perguntas de negócios por meio de indicadores de performance (KPIs), classificações e visualizações. O relatório a seguir apresentará análises baseadas em dados do banco BanVic, com foco em perguntas específicas sobre o banco.

2 Materiais

2.1 Dataset

O banco possui várias tabelas que armazenam dados essenciais para análise, fornecidas em formato csv. A tabela *agências* inclui informações sobre o código, nome e tipo. A tabela *clientes* contém dados como código, nome, data de nascimento, permitindo segmentar clientes por perfil demográfico. A tabela *colaborador_agencia* relaciona colaboradores com agências, enquanto *colaboradores* fornece informações sobre os empregados do banco, como cargo e data de admissão. A tabela *contas* registra dados sobre as contas bancárias, incluindo saldo. A tabela *propostas_credito* detalha informações sobre propostas de crédito, como valor e status. Por fim, a tabela *transacoes* contém dados sobre as transações bancárias, incluindo tipo e valor, sendo usada para identificar padrões e sazonalidade nas transações. Os dados precisaram de tratamento para que pudessem ser utilizados nas análises propostas neste relatório.

2.2 Tratamento dos Dados

Foram realizadas várias modificações nos dados para melhorar a análise. Em *contas*, foram criadas novas colunas como a *classificacao* para categorizar o saldo das contas em diferentes faixas, a *ano_abertura* e *ano_ultimo_lanc* para extrair os anos de abertura e do último lançamento, e a *t_inat*, que calcula o tempo de inatividade da conta. Em *transacoes*, foram extraídas as colunas *ano* e *mes* a partir da data das transações. Em *agencias*, foi calculado o número total de transações por agência. Além disso, foi removido o ano de 2023 dos dados de transações, pois estavam incompletos. Foi criado um mapeamento entre códigos de agência e seus nomes em *contas_por_agencia_df*. Essas mudanças visaram criar variáveis informativas, facilitar a análise temporal, agregar informações e melhorar a visualização, além de remover dados incompletos para evitar vieses, resultando em uma análise mais precisa e útil para o negócio bancário. Com os dados tratados, as análises podem ser realizadas. Neste contexto, algumas perguntas de negócio foram formuladas e os KPIs e visualizações correspondentes foram desenvolvidos para responder a estas perguntas.

2.3 Tecnologia

O tratamento dos dados foi realizado em linguagem Python, através do uso de notebook jupyter. Os gráficos foram plotados com a biblioteca Plotly em Python e os dashboards foram criados com Power BI. O notebook de desenvolvimento e os dashboards estão disponíveis no repositório no GitHub.

3 Perguntas de Negócio e KPIs

A análise de dados bancários depende de perguntas estratégicas que ajudam a entender melhor o comportamento dos clientes e a operação do banco. As perguntas formuladas abordam aspectos essenciais, como a distribuição de clientes

por saldo de conta, a segmentação entre clientes físicos e jurídicos, a preferência por tipos de agência (digital ou física), e a localização das agências com maior volume de clientes. Essas questões são fundamentais para identificar tendências de comportamento, otimizar os serviços oferecidos e aumentar a eficiência no atendimento. Alguns exemplos de perguntas que podem ser feitas são apresentados a seguir:

- Qual a distribuição dos clientes por saldo em conta
- Quais os tipos de clientes que possuem contas no banco (físico ou jurídico)?
- Considerando os tipos de agência (digital e física), qual tipo tem mais clientes?
- Baseado na localização, qual das agências tem mais clientes e por quê?
- Qual ano teve mais abertura de contas?
- Qual o tipo de transação mais frequente?
- Qual a frequência de transações?
- Baseado no saldo em conta, quais os tipos de clientes que possuem conta no banco?
- Média do valor das transações;
- Média de transações por dia;
- Quantas propostas estão ativas?
- Quantas contas foram abertas por cada colaborador?
- Clientes que não fazem transações a muito tempo;

Os KPIs (Key Performance Indicators) e as visualizações relacionadas têm um papel crucial na monitoração e avaliação do desempenho do banco. Indicadores como o número de clientes por tipo de agência ou o número de contas abertas por ano fornecem informações valiosas para a tomada de decisões estratégicas. A média do valor das transações e a frequência delas são importantes para entender o volume de operações realizadas, enquanto a análise da quantidade de propostas ativas e contas abertas por colaborador ajuda a medir a performance individual e o impacto dos colaboradores. Esses KPIs, juntamente com as visualizações apropriadas, permitem a extração de insights que guiam ações corretivas e melhorias no negócio bancário. Alguns exemplos de KPIs que podem ser utilizados para avaliação do desempenho do banco são:

- KPI: Número de clientes por tipo de agência;
- KPI: Número de clientes por agência;
- KPI: Número de contas abertas por ano;
- KPI: Número de transações por tipo;
- KPI: Número de transações por mês;
- KPI: Número de transações por trimestre;
- KPI: Número de transações por ano;
- KPI: Número de clientes físicos/jurídicos;
- Visualização: gráfico de barras por tipo;

Baseadas nestas perguntas citadas, algumas análises serão realizadas com base nos dados fornecidos a seguir.

4 Análises

A seguir, algumas das análises citadas anteriormente serão apresentadas a seguir neste relatório.

4.1 Análises Principais

4.1.1 Tendências Trimestrais

Algumas análises foram requisitadas referentes aos padrões dos dados fornecidos. Uma delas envolveu a média trimestral de transações e o volume movimentado. Para isso, a média geral foi calculada considerando todos os trimestres dos anos contidos no banco de dados, cujos resultados podem ser visualizados na Tabela 1. Observa-se que tanto o trimestre com o maior número de transações quanto o com o maior volume movimentado é o quarto. Conforme ilustrado na Figura 1, essa tendência se deve ao mês de dezembro de 2022, que apresenta um número de transações significativamente superior ao padrão observado nos demais meses do ano.

Trimestre	Número de Transferências	Volume Transferido
1	788.769231	7.110626e+05
2	845.461538	6.644091e+05
3	937.692308	9.016052e+05
4	2923.461538	2.163221e+06

Table 1: Média do Número de Transferências e Volume Transferido por Trimestre

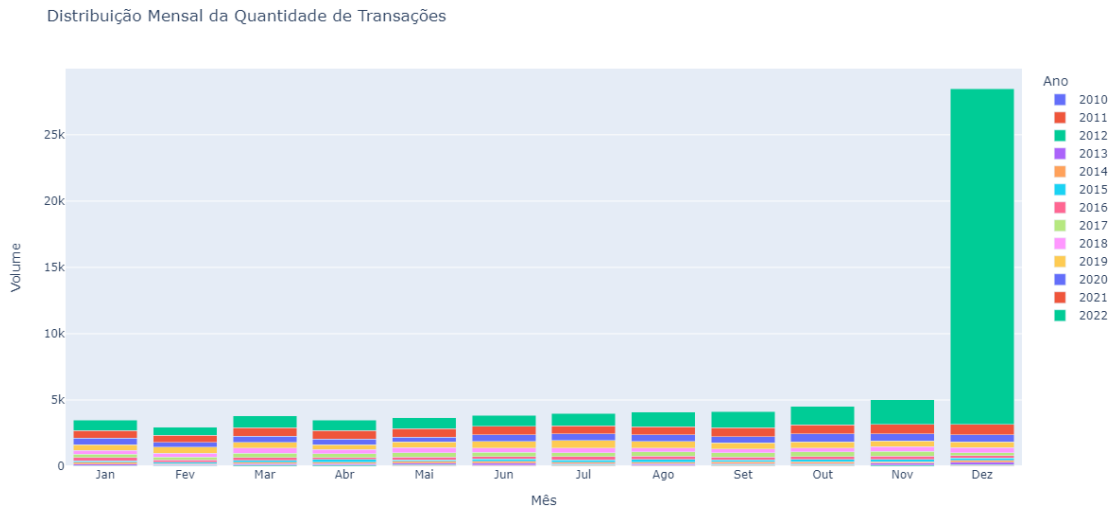


Figure 1: Distribuição de transações por mês e ano

4.1.2 Tendências Mensais

Essa análise foi motivada pela hipótese de um analista, que sugeriu que os meses que contêm a letra "R" em seu nome apresentam um padrão distinto na quantidade de transações em comparação aos meses que não possuem essa letra. Para verificar essa hipótese, os dados foram visualizados em um gráfico de barras, que exibe tanto o total de transações por mês quanto a distribuição desses totais ao longo dos anos 2. A partir dessas visualizações, conclui-se que não há relação entre a quantidade total de transações em um mês e a presença da letra "R" em seu nome. De forma geral, não é possível identificar um padrão na quantidade de transações. Além disso, no notebook, um teste estatístico foi realizado com os dados, e os resultados mostram que não há evidências para sustentar a hipótese.

4.2 Análises Complementares

4.2.1 Classificação dos Clientes

Uma análise importante que se pode fazer com relação aos dados fornecidos diz respeito à classificação dos clientes de acordo com o saldo disponível na conta. Para tal, foi utilizado o método de classificação baseado em quartis, que divide o conjunto de dados baseado em quatro grupos principais, e agrupa os dados dentro destes grupos. Qualquer valor que não se encaixe em uma destas quatro categorias é identificado como outlier no conjunto. Assim, dentro do contexto do saldo disponível dos clientes, o método baseado em quartis foi aplicado, e os limites foram calculados. Considerando a divisão de quartis, cinco categorias baseadas nos saldos das contas foram criadas para os dados, sendo estas:

- **Baixa:** saldo disponível no intervalo $[\text{Limite Inferior}, Q1)$.
- **Média-baixa:** saldo disponível no intervalo $[Q1, Q2)$.
- **Média:** saldo disponível no intervalo $[Q2, Q3)$.
- **Média-alta:** saldo disponível no intervalo $[Q3, \text{Limite Superior})$.
- **Alta:** saldo disponível a partir de Limite Superior.

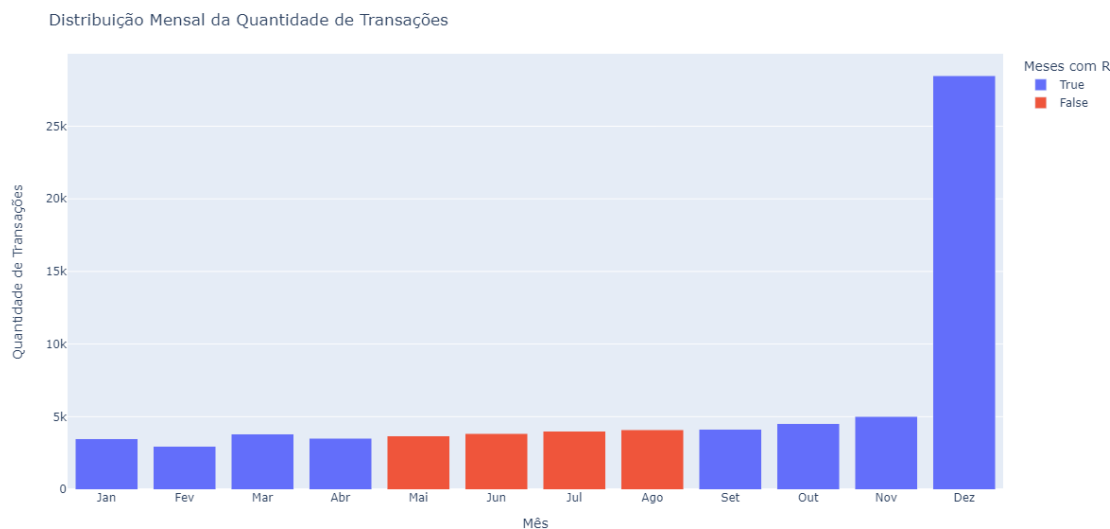


Figure 2: Visualização da distribuição mensal das transações

Segurança Financeira	Quantidade
baixa	250
média-baixa	250
média	249
média-alta	180
alta	70

Table 2: Distribuição dos clientes por segurança financeira

Estas categorias podem servir para classificar os clientes, por exemplo, por segurança financeira. Desta forma, é possível gerenciar melhor quais clientes podem receber propostas de crédito afim de mitigar os riscos de inadimplência. Neste contexto, a distribuição dos clientes no banco pode ser vista na Tabela 2.

4.2.2 Distribuição da Atividade nas Agências

A distribuição das contas no banco pode ser analisada considerando a quantidade de transações realizadas e o número de contas abertas por agência. Os gráficos 4 e 3 mostram que, apesar de ser uma das agências mais recentes, a Agência Digital possui o maior número de contas abertas e transações realizadas. Esse alto volume de novas contas pode ser atribuído à crescente facilidade de gerenciar contas bancárias e realizar transações virtualmente por meio de smartphones. Da mesma forma, o número elevado de transações no banco digital reflete a conveniência e agilidade oferecidas pelos serviços bancários online.

Diante desse cenário, é fundamental investir no aprimoramento da infraestrutura digital do banco. Medidas como a correção de bugs no aplicativo, o desenvolvimento de uma plataforma mais fluida e intuitiva e a ampliação do suporte ao cliente por meio de atendimento virtual são essenciais para melhorar a experiência do usuário. Além disso, estratégias de segurança robustas devem ser continuamente aprimoradas para garantir a confiabilidade das transações digitais. A implementação de tecnologias como inteligência artificial para suporte automatizado e análise de comportamento do usuário também pode contribuir para uma experiência mais personalizada e eficiente, reforçando a competitividade do banco no ambiente digital.

4.2.3 Atividade dos Clientes

Similar ao processo de classificação dos clientes, foi realizada uma classificação com respeito à atividade dos clientes. Com base nos dados fornecidos, foi identificado que há clientes que nunca realizaram transações no banco. Com processo similar, por divisão em quartis, foram criadas três categorias para os clientes, sendo estas Ativo, Moderadamente-ativo e Inativo. Os resultados encontrados podem ser visualizados na Tabela 3.

Distribuição das Contas por Agência

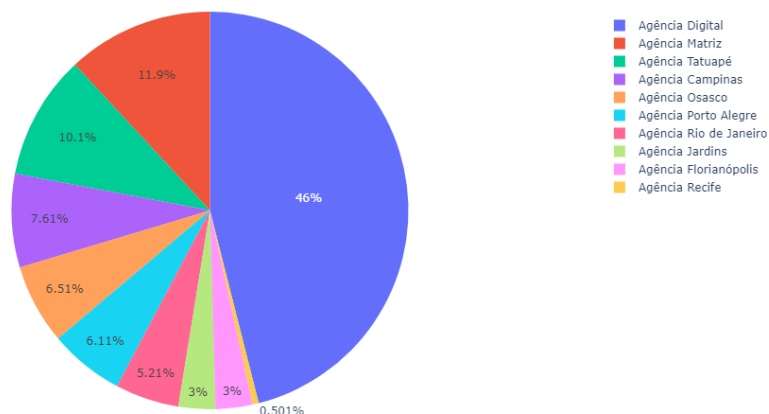


Figure 3: Distribuição das contas por agência

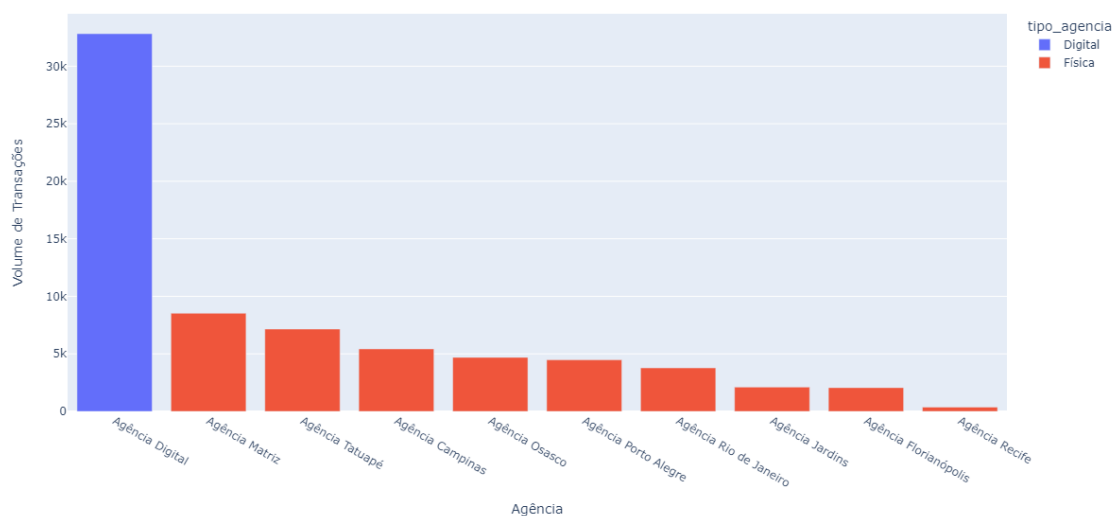


Figure 4: Volume de transações por agência

Estes resultados evidenciam que aproximadamente metade dos clientes do banco não fazem transações há um tempo considerável. Com estas informações é possível traçar estratégias para recuperação destes clientes, como contato direto e personalizado através de e-mails ou SMS, ofertas e benefícios exclusivos (e.g., isenção de taxas, condições especiais de crédito) e pesquisas de satisfação.

Inativo	486
Ativo	250
Moderadamente-ativo	250

Table 3: Resultados da classificação de atividade dos clientes

5 Fontes Externas

Ainda é possível comentar que tanto as visualizações quanto os KPIs dependem dos dados disponíveis para análise. Logo, uma análise completa requer cruzamento de dados de fontes distintas. Neste contexto, é possível aplicar diferentes fontes públicas para cruzar com os dados disponíveis e realizar análises mais profundas. A partir de pesquisa, algumas fontes de dados públicas foram coletadas e descritas na Tabela 4

Indicador	Fonte	Descrição	Aplicação
Índice Nacional de Preços ao Consumidor Amplo (IPCA)	IBGE	Mede a inflação oficial no Brasil, refletindo a variação dos preços de bens e serviços consumidos pela população.	Avaliar a correlação entre inflação e volume de transações; Analisar o impacto da inflação na inadimplência e comportamento de crédito; Melhorar a estratégia de precificação de produtos financeiros.
Renda Média e Distribuição de Renda	IBGE	Dados de renda per capita e distribuição de renda por região.	Segmentação de clientes com base no perfil socioeconômico; Mapeamento de regiões com maior potencial de expansão bancária; Avaliar a relação entre renda e taxa de inadimplência.
Taxa Selic e Taxas de Juros	Banco Central do Brasil (BCB)	A taxa Selic influencia diretamente os juros praticados pelos bancos e o comportamento de crédito da população.	Avaliar se clientes tomam mais empréstimos quando a taxa de juros está mais baixa.
Dados de Crédito e Inadimplência	Serasa Experian / SPC Brasil	Informações sobre consumidores inadimplentes e pontuação de crédito.	Identificar clientes em risco antes da inadimplência ocorrer; Personalizar ofertas de renegociação de dívida.

Table 4: Indicadores Econômicos e seu Impacto no Setor Bancário

Com a inclusão desses dados no Data Warehouse, o banco poderá tomar decisões estratégicas com maior embasamento. Na definição de estratégia de crédito, a relação entre o IPCA e o volume de transações permitirá ajustes nas taxas de juros e nas ofertas de crédito conforme o cenário econômico. Em relação à expansão comercial, a análise dos dados de renda e inadimplência ajudará o banco a identificar as regiões mais promissoras para ampliar suas operações. Além disso, na otimização de produtos bancários, caso a confiança do consumidor esteja baixa, o banco poderá direcionar seus esforços para produtos voltados à segurança financeira, como investimentos mais conservadores.

Seguindo um dos exemplos de aplicação fornecidos, pode-se comparar o IPCA com os dados de volume e quantidade de transações para verificar uma possível relação entre essas variáveis. Para isso, é calculado o coeficiente de correlação de Pearson, que mede a existência e a intensidade de uma relação linear entre duas variáveis. A interpretação do valor obtido para esse coeficiente pode ser visualizada abaixo:

Valor de r	Interpretação
$ \pm 1 $	Relação linear perfeita
$> \pm 0.70 $	Relação linear forte
$> \pm 0.50 $	Relação linear moderada
$> \pm 0.30 $	Relação linear fraca
0	Ausência de relação linear

Table 5: Interpretação do coeficiente de correlação r

Considerando o conjunto de dados fornecido, os resultados foram calculados para dois intervalos: (1) de 2010 a 2021 e (2) de 2010 a 2022. A necessidade de dois cálculos distintos deve-se à mudança no padrão das transações em 2022, mais especificamente devido ao mês de dezembro desse ano, que se configura como um outlier no conjunto de dados. Os valores dos coeficientes para ambos os intervalos são apresentados na Tabela 6. A partir da interpretação desses resultados, conclui-se que existe uma relação moderada entre o valor transferido anualmente e o IPCA anual, enquanto a relação entre o IPCA e o número de transferências é fraca.

Essas relações podem ser visualizadas nas séries temporais do número de transações, volume das transações e IPCA, apresentadas na Figura 5. Como as escalas dos valores diferem significativamente, foi necessário normalizar as três

Métrica	Intervalo 1	Intervalo 2
Valor Transferido	-0.5462	0.0891
Número de Transferências	-0.3295	-0.0191

Table 6: Coeficientes de correlação de Pearson para os diferentes intervalos

métricas utilizando o Z-score. No gráfico, observa-se que o valor transferido tende a aumentar quando o IPCA está baixo e a diminuir quando o IPCA está alto. Por outro lado, o número de transações apresenta um crescimento constante, independentemente do valor do IPCA, pois é um fator que está diretamente relacionado com o crescimento do número de contas no banco e com o surgimento do Pix em 2020. Assim, ambas as análises corroboram os resultados obtidos por meio da correlação de Pearson.

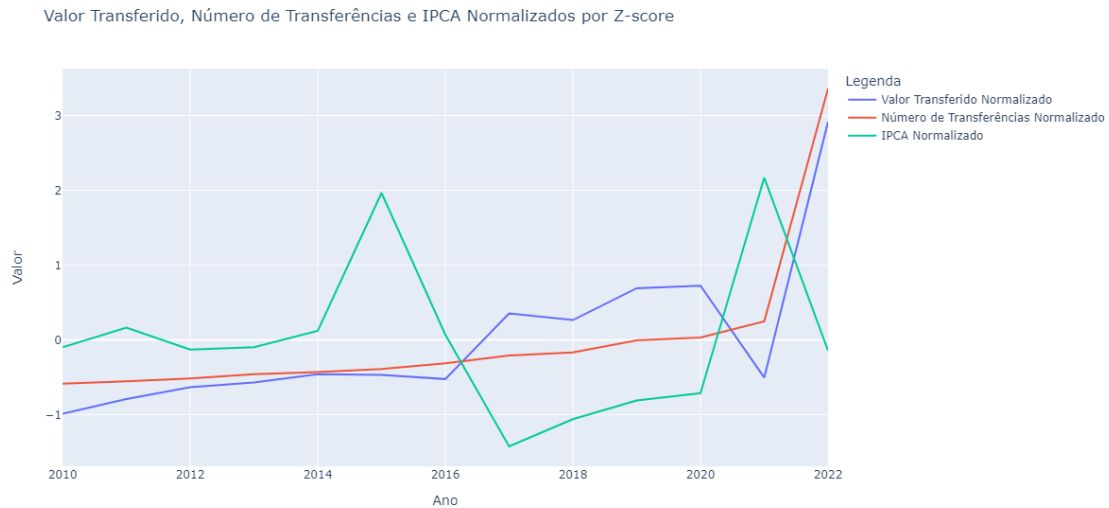


Figure 5: Séries temporais do número de transações, volume das transações e IPCA no intervalo de 2010 a 2022