

# Projeto de análise de Dados: Desafio BanVic

Jennifer Marques

7 de setembro de 2025

## 1 Introdução

Este relatório apresenta a análise dos dados fornecidos no desafio, descrevendo os passos adotados, os resultados alcançados e as principais conclusões obtidas.

## 2 Metodologia e Tratamento dos Dados

O processo de análise seguiu um fluxo estruturado de Extração, Transformação e Carregamento (ETL) para garantir a qualidade e a integridade dos dados.

A primeira etapa, de **Extração**, envolveu o carregamento de todos os arquivos CSV para o ambiente de análise utilizando a biblioteca Pandas da linguagem Python. Na fase de **Transformação**, foi realizado um tratamento dos dados, que incluiu:

- A conversão de colunas de data, que estavam em formato de texto (object), para o formato específico de data e hora (datetime), assim, possibilitando que as análises que exigiam esses dados fossem feitas com êxito.
- A criação de novas colunas (feature engineering), como dia da semana e nome do mês, para enriquecer as possibilidades de análise.
- A junção sequencial das tabelas de clientes, contas, agências e transações através da operação de *merge*, resultando em um único DataFrame que serviu como base analítica central.
- O enriquecimento do conjunto de dados com uma fonte externa – a cotação diária do dólar, obtida via API do Banco Central do Brasil.

A etapa de **Carga** consistiu na exportação dos dados limpos e formatados para a ferramenta de visualização Google Looker Studio, onde o dashboard interativo foi feito.

**Observação técnica:** As principais transformações foram realizadas em Python usando **pandas**, com verificações de integridade (contagem antes/depois, checagem de duplicados, e tipos).

## 3 Análise dos Dados e Resultados

### 3.1 Questões para área Comercial (resumo)

Para demonstrar o valor da análise de dados à diretoria comercial, a análise foca em traduzir dados brutos em respostas para perguntas de negócio críticas. A proposta para convencer a diretora é ir além do marketing tradicional, respondendo a perguntas como:

- **Quem são nossos clientes de maior valor e onde se concentram?** Para responder, seriam utilizados os KPIs de *Saldo Médio por Cliente*, *Volume Total Transacionado* e a *Distribuição Geográfica*. A própria análise de agências (detalhada na Seção 3.4) já fornece um ponto crítico: a "Agência Digital" concentra um volume de transações muito maior que as agências físicas, sugerindo que o cliente de maior atividade para o BanVic já é, por natureza, um cliente digital.
- **Quais canais geram maior ticket médio?** Utilizando o KPI de *Tipo de Transação Mais Comum (PIX, Saque, Débito)*, seria possível segmentar o comportamento de clientes. Uma análise preliminar dos dados agregados já sugere, por exemplo, uma forte predominância do PIX como principal meio transacional, refletindo uma tendência de mercado que o BanVic pode explorar para produtos digitais. Essa análise permitiria entender se o uso de canais mais tradicionais, como o saque, está concentrado em um perfil específico de cliente, possibilitando ações de marketing e educação financeira direcionadas.

Essa abordagem permite que a área comercial otimize investimentos, personalize a comunicação e atue de forma proativa na retenção de clientes valiosos, provando que a análise de dados é uma ferramenta poderosa para o crescimento do negócio.

### 3.2 Dimensão de datas: dias da semana e meses pares/ímpares

Após as análises sobre a distribuição das transações nos dias da semana os resultados mostram que a quinta-feira é o dia com maior concentração de operações no banco. O volume financeiro neste dia (R\$ 6,64 milhões) é aproximadamente 72% maior que o da sexta-feira (R\$ 3,87 milhões), o segundo dia mais movimentado, evidenciando um padrão de altíssima concentração. De forma similar, o número de transações na quinta-feira (21.120) chega a ser quase o triplo da média dos outros dias úteis (aprox. 7.500 transações). Este padrão pode estar associados a ciclos de pagamento de salários e fornecedores, representando um dia de alta liquidez para os clientes. Este pico representa não apenas uma oportunidade comercial, mas também um ponto de estresse operacional para os sistemas do BanVic, exigindo máxima infraestrutura de TI para garantir a estabilidade.

Dia da Semana	Total de Transações	Volume Total Movimentado (R\$)
Segunda-feira	7.176	3.018.450,16
Terça-feira	7.430	2.232.560,21
Quarta-feira	7.883	3.589.484,61
<b>Quinta-feira</b>	<b>21.120</b>	<b>6.646.010,94</b>
Sexta-feira	13.571	3.870.517,83
Sábado	7.380	2.790.520,53
Domingo	7.439	3.557.910,12

Tabela 1: Transações e volume por dia da semana.

**Meses pares vs ímpares** Foi testada a hipótese de que meses pares teriam um volume médio de transações significativamente maior. A análise não apenas refutou esta sugestão, como revelou o padrão oposto: o valor médio das transações em meses ímpares foi de R\$ 513,38, um montante aproximadamente 86% superior à média registrada em meses pares, que foi de R\$ 275,62. Este resultado destaca a importância crucial de validar hipóteses com dados, evitando que decisões sejam tomadas com base em premissas erradas. Para aprofundar a análise temporal, propõem-se duas novas investigações:

- **Análise de Início vs. Fim de Mês:** Verificar se existe um pico de saques e depósitos nos primeiros dias úteis do mês (efeito "dia do pagamento"), para otimizar a gestão nas agências.
- **Análise de Sazonalidade:** Comparar o volume de compras em meses com datas comemorativas (como Dezembro) com a média dos outros meses, para planejar campanhas de marketing temáticas.

### 3.3 Correlação com cotação do dólar

Foi investigada a correlação entre a cotação diária do dólar e o valor das transações dos clientes do BanVic. Conforme o gráfico de dispersão abaixo, não foi observada uma relação linear clara entre as duas variáveis.

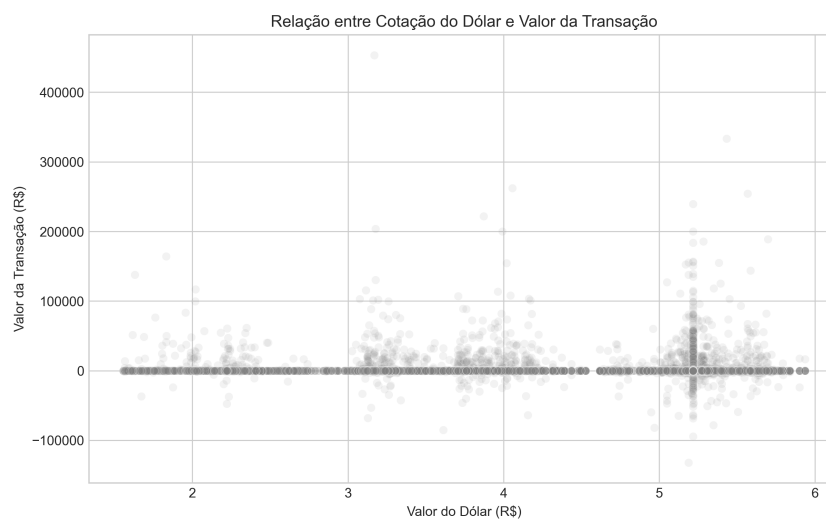


Figura 1: Relação entre a Cotação do Dólar e o Valor das Transações.

Foi calculado o coeficiente de Pearson entre cotação diária do dólar e valor das transações agregadas (por dia), com seu valor próximo a zero (-0.0299), confirmando a impressão visual do gráfico. Isso indica a ausência de correlação linear com a moeda estrangeira, o que não impacta diretamente no comportamento transacional da base de clientes analisada. Este achado, embora nulo, é estrategicamente valioso: ele facilita uma vez que a gestão do BanVic passa a focar em fatores mais diretamente ligados à economia local (como dados de emprego), em vez de se preocupar com a volatilidade do câmbio para a sua base de clientes atual.

Para futuras análises, recomenda-se o enriquecimento do Data Warehouse do BanVic com as seguintes fontes de dados públicos:

- **IBGE Cidades:** Dados de PIB per capita e população para correlacionar o desempenho das agências com o potencial econômico de suas localidades e responder perguntas como: "O baixo desempenho da Agência Recife é justificado pelo menor potencial econômico da região ou há um problema de gestão local?"
- **Banco Central (BCB):** Séries históricas da taxa SELIC para entender como a política monetária afeta o comportamento de crédito e investimento dos clientes, permitindo ao banco modelar o impacto de futuras altas de juros na carteira de crédito e se antecipar a possíveis aumentos na inadimplência.

### 3.4 Ranking de agências (últimos 6 meses)

A análise do total de transações por agência nos últimos seis meses revelou uma performance significativamente heterogênea. A Agência Digital, por exemplo, registrou um total de transações mais de 40 vezes superior ao da Agência Recife (16.806 contra 374), indicando realidades operacionais e de mercado drasticamente distintas entre as unidades. As tabelas a seguir apresentam os rankings das três agências com maior e menor volume de transações, respectivamente.

Nome da Agência	Total de Transações
Agência Digital	16.806
Agência Matriz	2.345
Agência Tatuapé	2.246

(a) Ranking das 3 Melhores Agências.

Nome da Agência	Total de Transações
Ag. Florianópolis	1.524
Agência Jardins	1.121
Agência Recife	374

(b) Ranking das 3 Piores Agências.

Tabela 2: Comparativo de desempenho das agências por total de transações.

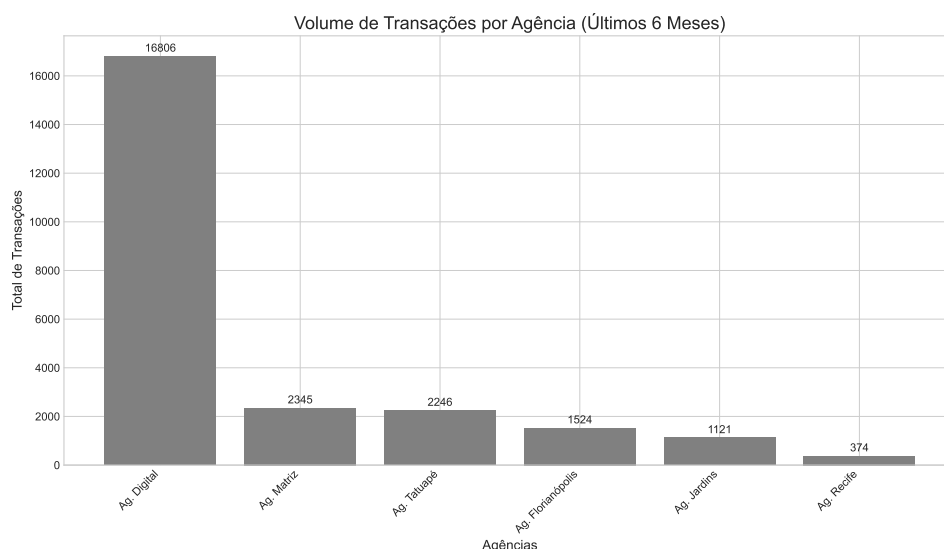


Figura 2: Disparidade no total de transações entre as agências

A disparidade observada entre as agências de melhor e pior desempenho é um ponto de atenção crucial. Essa informação permite que a gestão direcione esforços para investigar as causas do baixo volume nas unidades de menor performance, ao mesmo tempo em que pode identificar e replicar as boas práticas das agências de melhor performance. As hipóteses para essa disparidade podem variar desde fatores demográficos e econômicos locais até a eficácia das equipes comerciais ou o perfil de adoção de produtos digitais pelos clientes de cada regional.

## 4 Dashboard e KPIs principais

Para consolidar os principais indicadores e permitir a exploração interativa dos dados, foi desenvolvido um dashboard no Google Looker Studio. O painel apresenta os KPIs centrais e permite a filtragem dinâmica por período, agência e cliente.

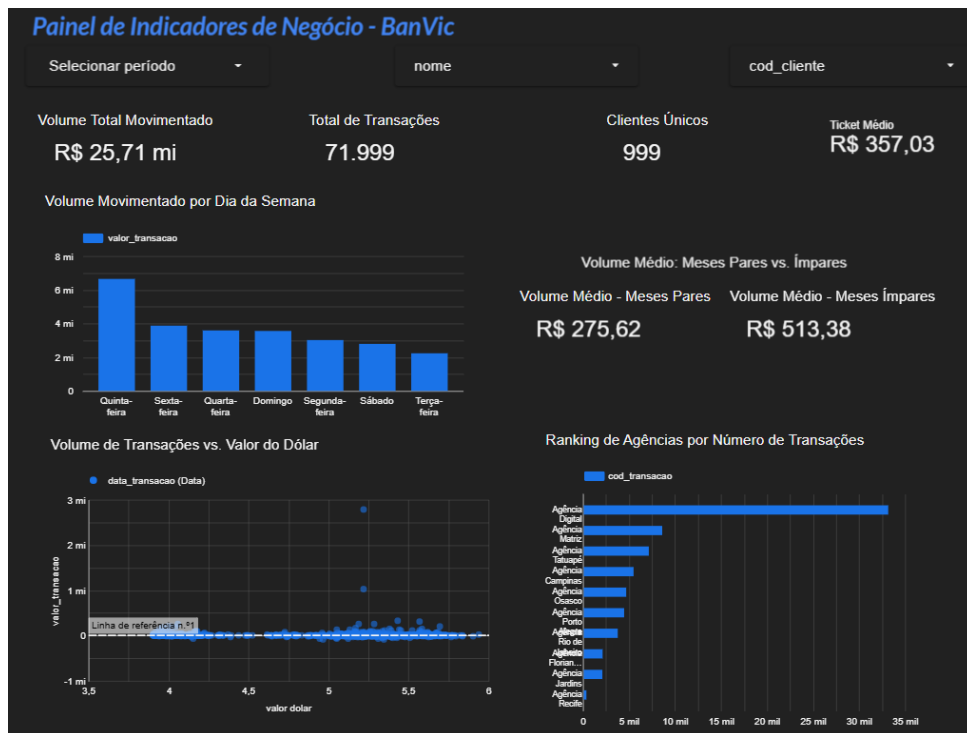


Figura 3: Dashboard com os Principais Indicadores de Negócio do BanVic.

O dashboard foi projetado para fornecer à gestão uma visão rápida e clara da performance do negócio.

KPIs principais:

- **Volume Total Transacionado** - Soma do valor de todas as transações em um período. Monitora o crescimento da movimentação financeira e identifica padrões sazonais de liquidez.
- **Número de Transações** - Quantidade de operações realizadas por clientes em um período. Acompanha o nível de engajamento dos clientes e a utilização dos canais de atendimento.
- **Total de Clientes Únicos** - Quantidade de clientes únicos por agencia.
- **Ticket Médio por Cliente** - Média de valor movimentado por cliente ativo. Identifica clientes de maior valor para embasar estratégias de segmentação e retenção.

Ele pode ser acessado através do seguinte link: [Dashboard BanVic KPIs](#).

## 5 Ferramentas Utilizadas

A execução deste projeto foi realizada pela utilização de ferramentas padrão de mercado para análise de dados:

- **Linguagem Python:** Escolhida pela versatilidade e pela grande disponibilidade de bibliotecas para análise de dados.
- **Pandas & NumPy:** Bibliotecas utilizadas para a manipulação, limpeza, tratamento e agregação dos dados.
- **Matplotlib & Seaborn:** Bibliotecas empregadas para a criação das visualizações estáticas presentes neste relatório.

- **Jupyter Notebook:** Serviu como ambiente de desenvolvimento interativo para a exploração e documentação do processo de análise.
- **Google Looker Studio:** Ferramenta utilizada para a criação do dashboard interativo devido à sua facilidade de uso e integração com fontes de dados CSV.
- **Overleaf (LaTeX):** Escolhido para a elaboração deste relatório, visando um resultado final profissional e bem estruturado.

## 6 Conclusão e Recomendações Estratégicas

Este projeto de análise de dados demonstrou o imenso potencial contido nos dados transacionais do BanVic. As principais descobertas incluem a notável disparidade de performance entre as agências e a forte concentração de atividade financeira nas quintas-feiras, indicando padrões de comportamento claros que podem ser explorados estrategicamente.

Com base nos insights gerados, as seguintes recomendações são propostas:

1. **Implementar um Programa de Performance de Agências:** Iniciar um projeto para investigar as causas do baixo desempenho das agências com piores rankings. Ao mesmo tempo, documentar e disseminar as práticas operacionais e comerciais das agências de topo como um modelo para as demais.
2. **Capitalizar no "Efeito Quinta-feira":** Desenvolver ações de marketing e vendas direcionadas para as quintas-feiras, aproveitando o pico de atividade dos clientes. Isso pode incluir desde a garantia de máxima performance dos sistemas até a oferta de produtos de crédito pré-aprovado neste dia.
3. **Enriquecer o Data Warehouse:** Adotar a prática de integrar fontes de dados externos (como IBGE e BCB) ao ambiente de dados do BanVic para enriquecer futuras análises e permitir um planejamento estratégico ainda mais eficiente.

As descobertas apresentadas de forma detalhada neste relatório e de forma interativa no dashboard, não apenas fornecem respostas, mas provam o valor de uma cultura analítica que substitui percepções por fatos.