

Relatório – Desafio Análise de Dados/Engenharia de Analytics

1. Contexto

O Banco Victória S.A, por meio de iniciativa de sua CEO, demonstrou interesse em aprofundar o entendimento do comportamento de seus clientes visando melhorar a prestação de serviços. A equipe técnica do banco reportou que há intenção de implementar técnicas mais avançadas para a análise de dados, visto que atualmente muitas tarefas são feitas manualmente. Portanto, há necessidade de otimizar as operações internas do banco em relação aos dados.

No entanto há resistência por parte da Diretora Comercial em perceber o valor de um projeto de analytics e o investimento em uma infraestrutura moderna de dados.

2. Objetivo

Construir uma infraestrutura moderna de dados, garantindo informações confiáveis e flexíveis para análises, possibilitando que as lideranças do Banco Victória S.A. tomem decisões mais assertivas em relação a investimentos estratégicos, marketing e expansão.

3. Solução Proposta – Processos e transformação de dados

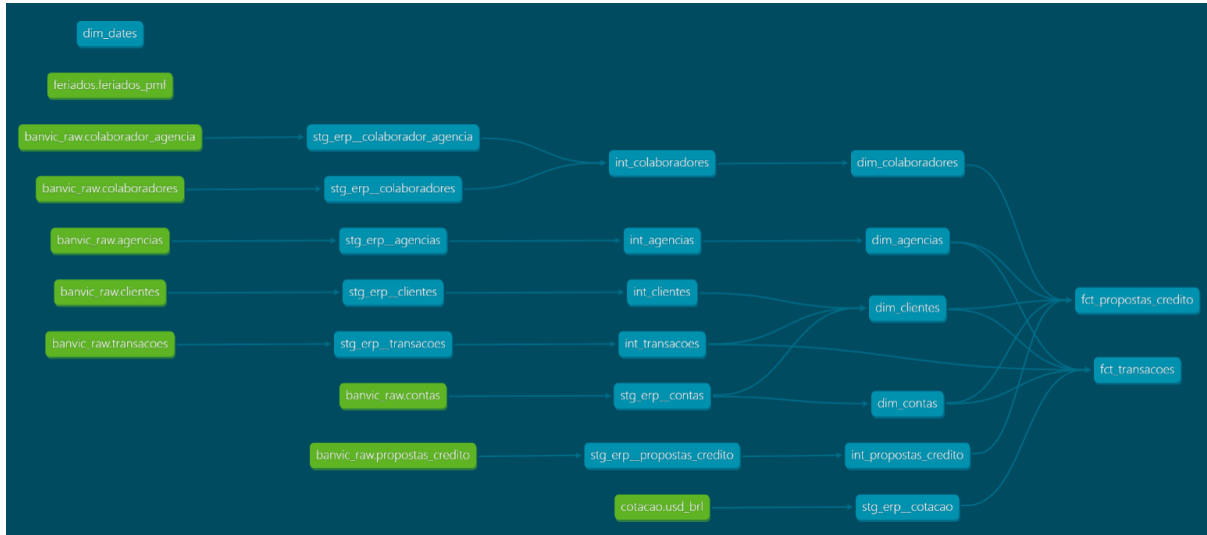
Itens B e F

- Ingestão de dados
 - Considerando que trata-se de um cenário onde os dados já estariam disponibilizados em um ambiente para serem trabalhados de forma analítica, não foi realizado nenhum processo de ingestão. Porém, uma vez que deseja-se um ambiente de transformação compatível com SQL para aplicar-se o framework de modelagem imposto pelo dbt, foi utilizada a técnica de *seeding* para carregar os dados disponibilizados (CSV) no banco de dados.
Recomenda-se a utilização de um sistema OLAP (Databricks, Snowflake, Redshift, etc), para implementação produtiva da solução proposta, contudo, foi utilizado um banco OLTP (PostgreSQL) para simular um *data warehouse* em ambiente local, o que facilita a reprodutibilidade do projeto como um todo, sendo disponibilizado junto do projeto as configurações necessárias para implantar uma instância do banco em ambiente containerizado. Adicionalmente, o uso do dbt como abstração da camada de processamento, torna a migração do código de transformação entre ambientes de processamento muito mais fácil, sendo necessário apenas alterar as funções específicas do Postgres utilizadas de forma explícita no código de transformação.
- Modelagem de dados:
Como mencionado anteriormente, para a modelagem de dados foi adotado o framework proposto pelo dbt. Seguindo a metodologia de Kimball (star schema) e analisando as tabelas apresentadas, chega-se nas seguintes tabelas fato e dimensão:
 - Fato transações e propostas de crédito
 - Dimensões: clientes, agências, contas, datas, colaboradores



O diagrama conceitual pode ser visualizado acessando o [link](#).

- Seguindo a filosofia proposta pelo dbt, a camada de staging mapeia as tabelas fonte 1:1, realizando operações de *type casting*, renomeação consistente e seleção de colunas. Já a camada intermediária realiza algumas transformações e joins e as marts implementam as tabelas fato e dimensões finais. Todas implementadas utilizando o dbt, realizando:
 - Validações de integridade (left joins entre fatos e dimensões)
 - testes de dados (unicidade, não nulidade, etc) para garantir a qualidade dos dados entregues
 - documentações em YAML, trazendo informações para fundamentar as decisões de modelagem, necessidades do negócio e quaisquer outras particularidades.



- Integração de dados externos (cotação do dólar, feriados nacionais, taxa Selic) para análises enriquecidas.
- Para a camada de consumo, foram construídas visualizações servidas em Power BI, trazendo KPIs estratégicos e gráficos intuitivos para auxiliar na tomada de decisões/entendimento do negócio.

Em suma, as ferramentas usadas nesse processo foram:

- Desenvolvimento: VSCode, SQLFluff, yamllint, uv
- Data Warehouse: Docker, PostgreSQL
- Modelagem: dbt
- Visualização: PowerBI

4. Achados e Observações

Item a-1

O *Business Intelligence* (BI) é importante na gestão de negócios, independente do seguimento. Apoia a tomada de decisões estratégicas e permite a alocação eficiente de recursos nos canais com maior retorno sobre investimento. No caso específico reportado pela Diretoria Comercial nesse

desafio, o BI pode se tornar fundamental para viabilizar segmentações mais precisas, compreender melhor o comportamento dos clientes e planejar campanhas mais eficazes.

Os dados disponíveis eram limitados, visto que se trata de um projeto piloto proposto por Lucas Johnson e inclui apenas uma parte dos dados do BanVic. No entanto, foi possível fazer um exemplo de pipeline de dados e análises que poderão ajudar a responder perguntas estratégicas.

Algumas perguntas de negócio que podem ser feitas incluem:

1. Quais agências concentram maior número de transações e aprovação de crédito?
2. Quando (dia/mês) o banco tem maior volume e valor de transação?
3. Gragalos: há horários/dias com taxas de transação anormais?
4. Sazonalidade: o volume e o valor de transações mudam de acordo com o período do ano/semana? (meses específicos, feriados)

Item a-2

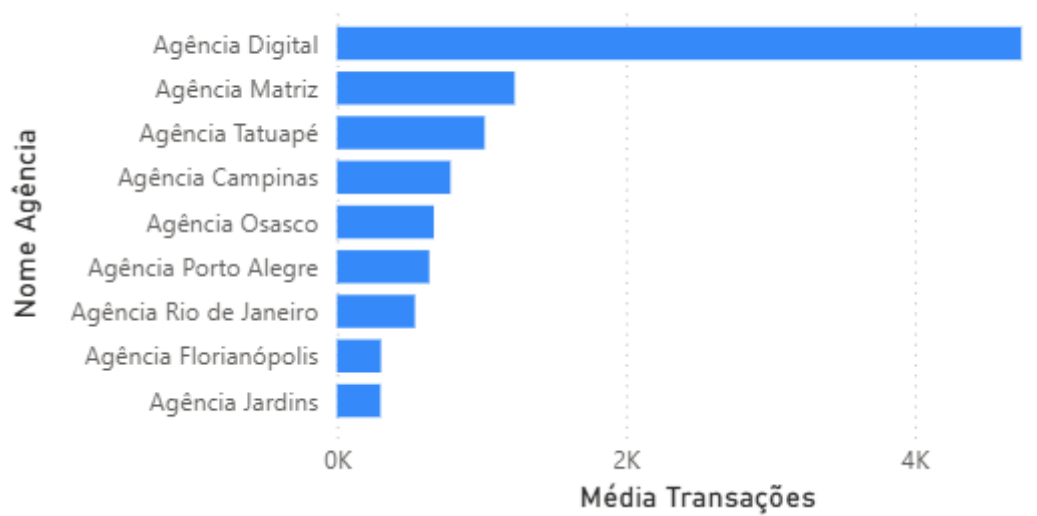
Os KPI's que podem responder essas e mais perguntas podem ser construídos utilizando o Power BI. Foram calculados:

- Volume total de transações (número de transações feitas);
- Valor movimentado (Valor das transações) total e média;
- Número total de clientes;
- Número total de propostas de crédito;
- Taxa de aprovação de propostas de crédito.

Adicionando filtros podemos aprofundar essas análises, identificando o comportamento desses KPI's ao longo do tempo, por cidade, por agência, por tempo de conta, por tipo de conta, e outros filtros incluídos no Dashboard do Power BI.

Na análise exploratória inicial podemos observar no gráfico abaixo que a Agência Digital concentra volume de transações significativamente maior em comparação com as agências físicas.

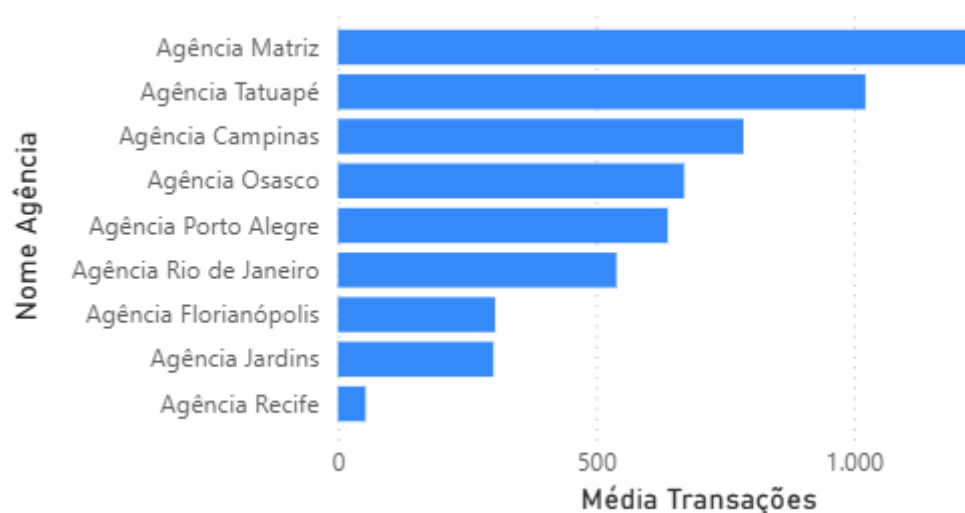
Média de transações por agência



A Diretora Comercial considerava investir em marketing em cidades com mais agências. Entretanto, a análise do gráfico acima levanta a questão: qual o retorno real de investir apenas em agências físicas?

Ainda em relação à performance das agências, quando a agência digital é excluída das análises a distribuição de transações ainda parece ser heterogênea.

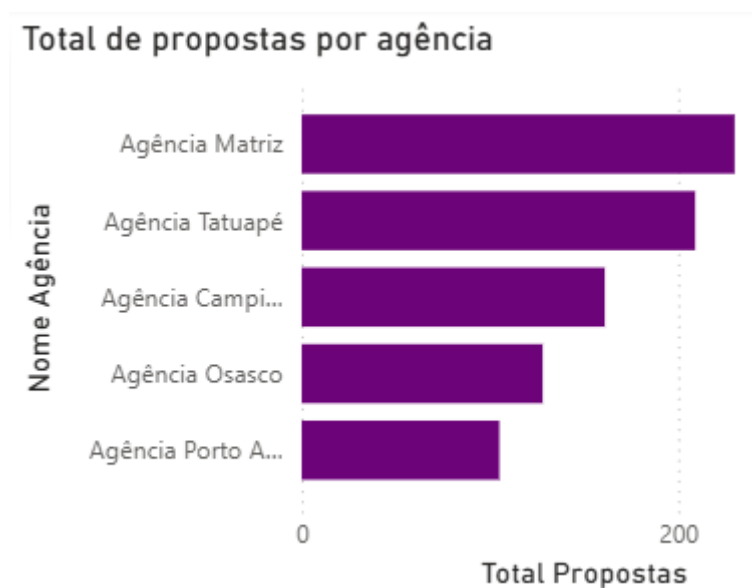
Média de transações por agência



Esse comportamento se mantém quando avaliamos as performances de crédito, com a inclusão da agência digital.



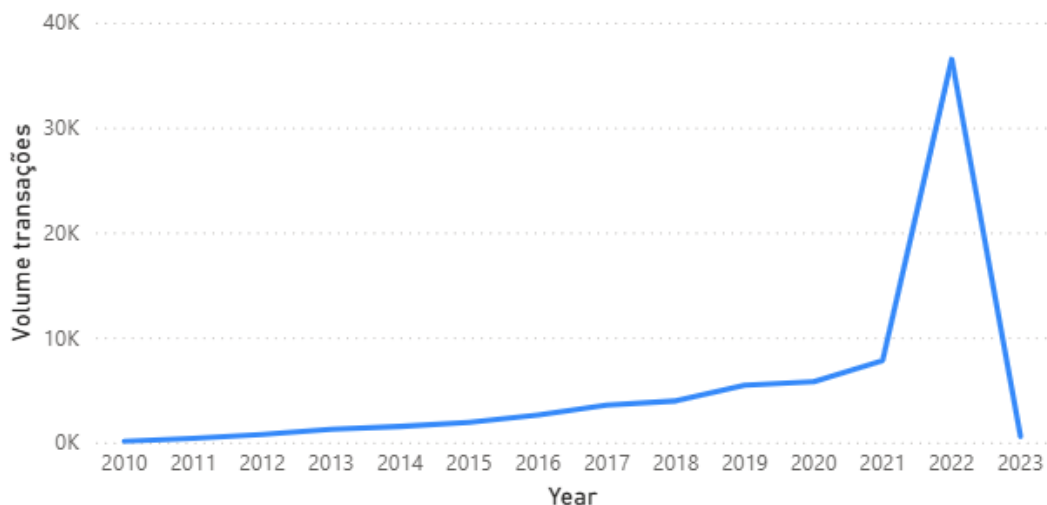
E sem a agência digital:



Ainda, houve um pico acentuado de transações no final de 2022, indicando alguma ocorrência. Podendo ser um fenômeno nacional ou mundial causando um distúrbio econômico, um problema com as extrações/inputações de dados, ou até mesmo alguma fraude em larga escala. Uma análise mais profunda é necessária para entender o que aconteceu, superficialmente é possível dizer que o número de clientes aumentou em 36% mas o volume de transações e o valor total transacionado aumentaram em uma ordem muito maior (cerca de 400%), contudo, a ordem do aumento é semelhante entre o volume e o valor

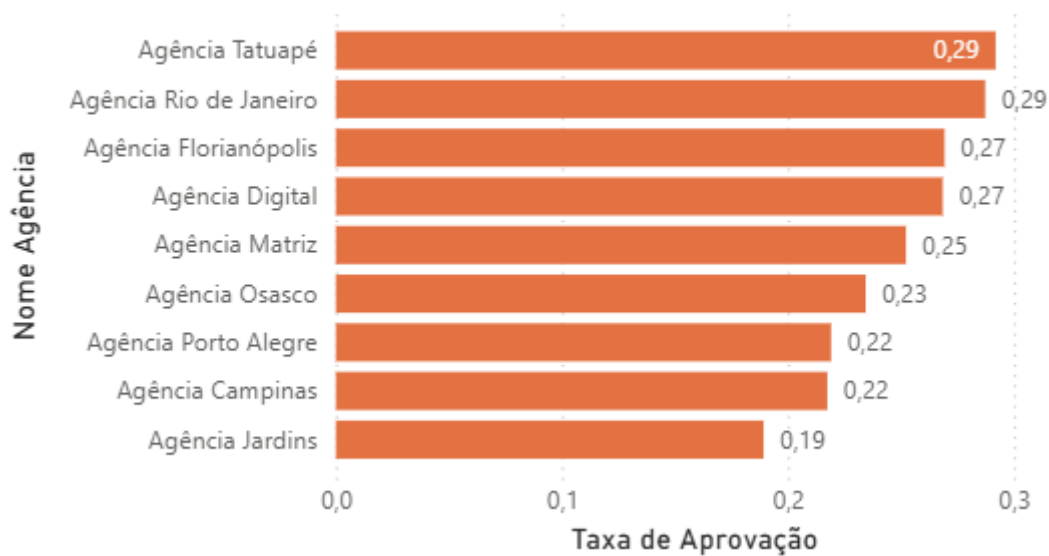
transacionado.

Volume transações por ano



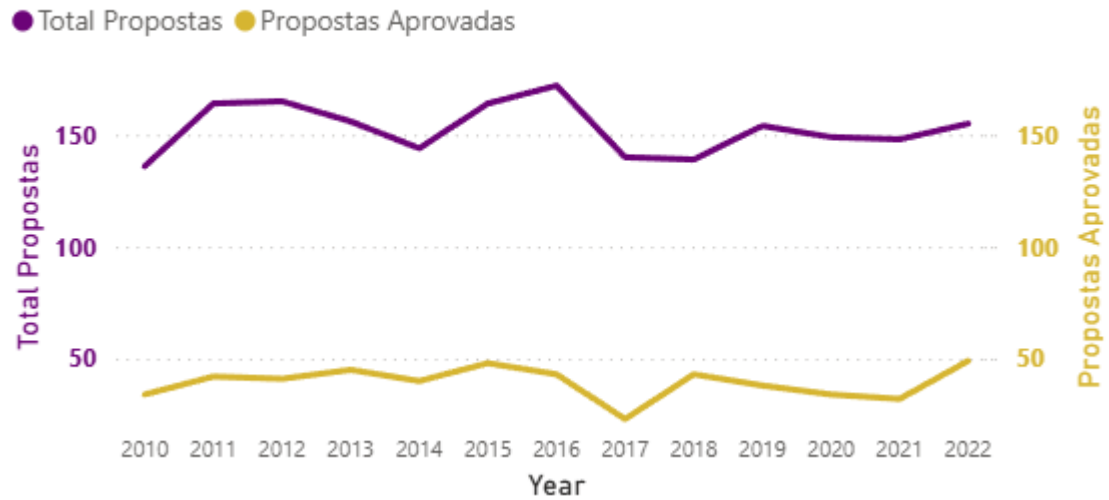
A taxa de aprovação de propostas de crédito foi de 27,5%, sendo a Agência de Tatuapé/SP a agência com melhor taxa.

Taxa de aprovação (%) por agência



O total de propostas feitas é maior que o total de propostas aprovadas e o comportamento de ambas parece acompanhar o mesmo padrão ao longo dos anos.

Total de propostas e propostas aprovadas por ano

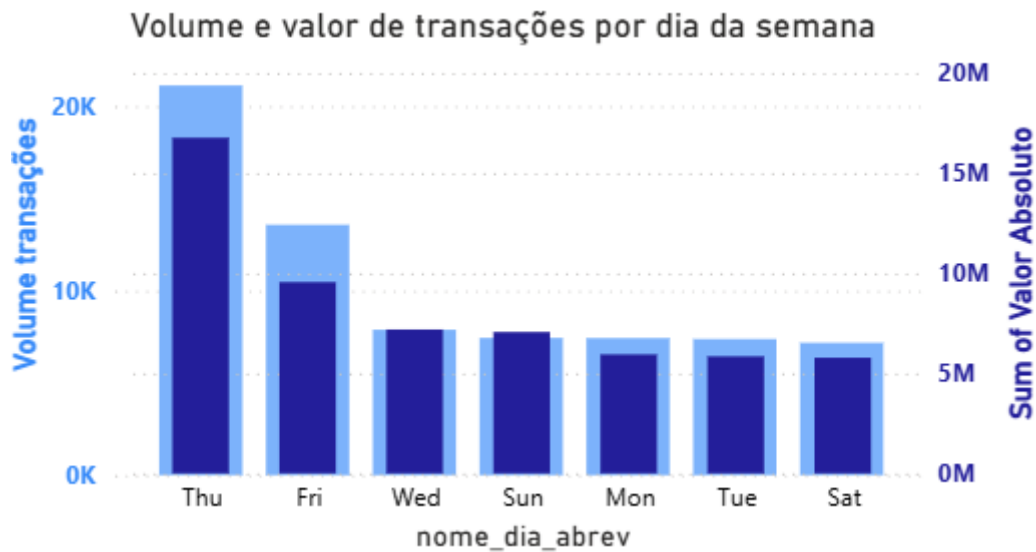


♦ Obs: Qualidade dos Dados

- Foram encontrados valores nulos em campos críticos (ex.: cliente id 528 com propostas de crédito e transações efetuadas não existe na tabela de clientes).
- Necessidade de atuação conjunta com a equipe de Engenharia de Dados para avaliar falhas no pipeline de ingestão ou verificação dos dados na fonte pela equipe do BanVic.

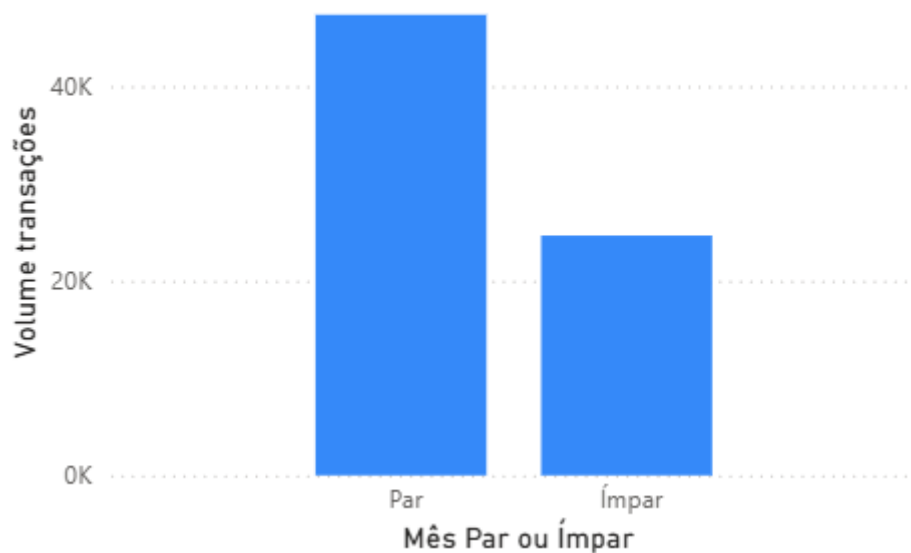
Item a-3

Utilizando uma dimensão de datas completa e bem dimensionada, podemos identificar com o gráfico abaixo que o dia da semana em que o BanVic tem maior número e valor de transações é a quinta-feira.



Um analista do BanVic sugeriu que meses pares têm maior volume médio de transações em comparação com meses ímpares. Visualmente essa hipótese se confirma, no entanto é necessário fazer uma análise estatística de teste de hipótese e possivelmente análises de regressão para controlar os resultados por fatores confundidores para determinar se esse resultado é aleatório ou tem alguma causa específica.

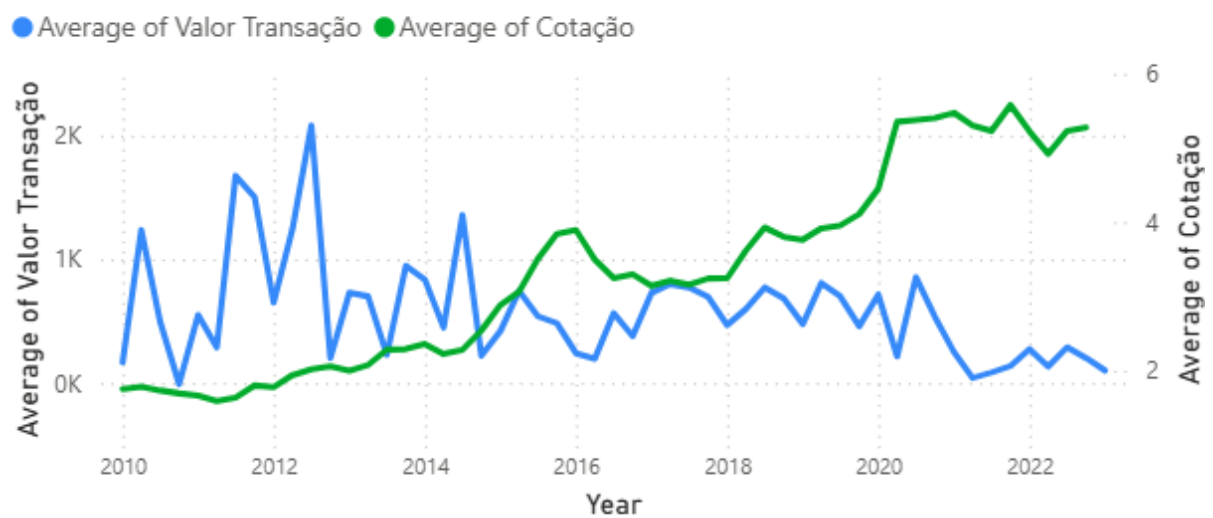
Volume transações by Mês Par ou Ímpar



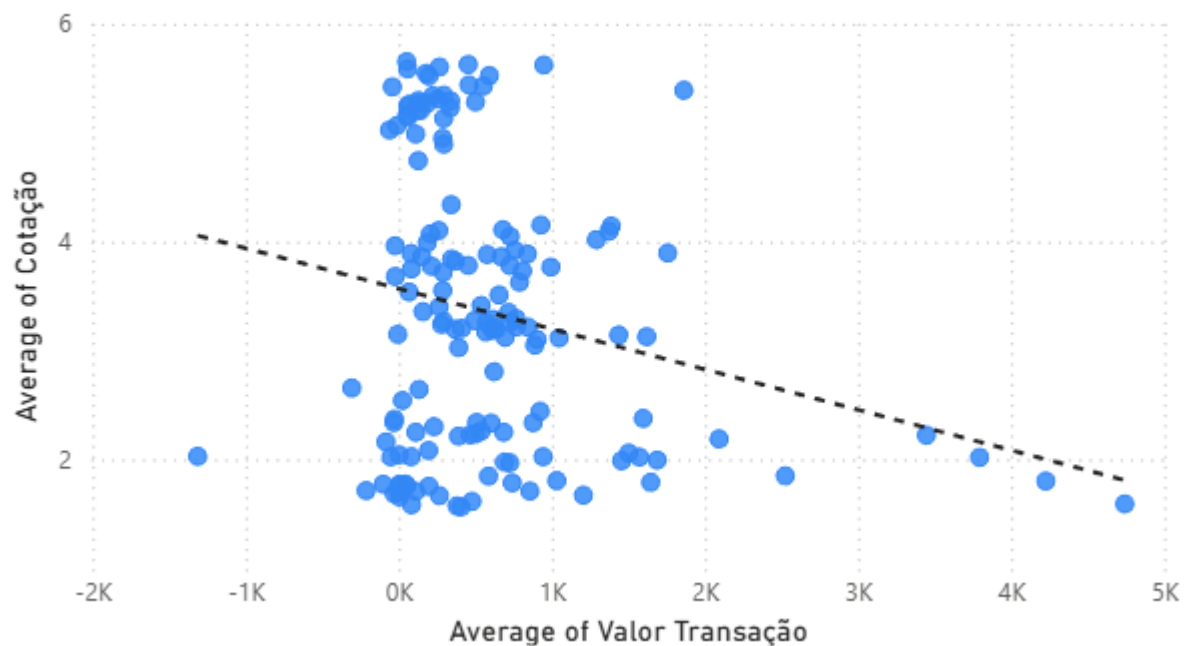
Item a - 4

Um especialista de tecnologia do BanVic solicitou uma proposta de dados externos para enriquecer as análises dos dados. Um exemplo utilizado nesse projeto foi a cotação do dólar. Houve correlação inversa com o valor médio das transações (observada em análise temporal, mas ainda sem teste estatístico de significância).

Valor de transações e cotação do dólar por ano



Valor movimentado por cotação do dólar por ano

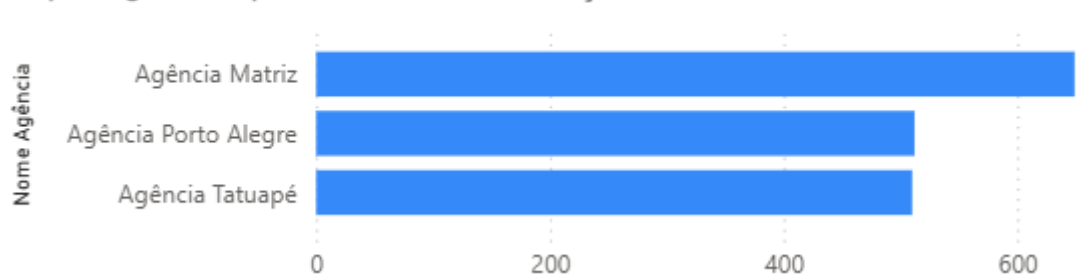


- Outros dados públicos recomendados:
 - Feriados nacionais → impacto direto em volume de transações;
 - Dados geográficos → impacto em expansão estratégica;
 - Taxa Selic → relação com crédito e taxas de juros.

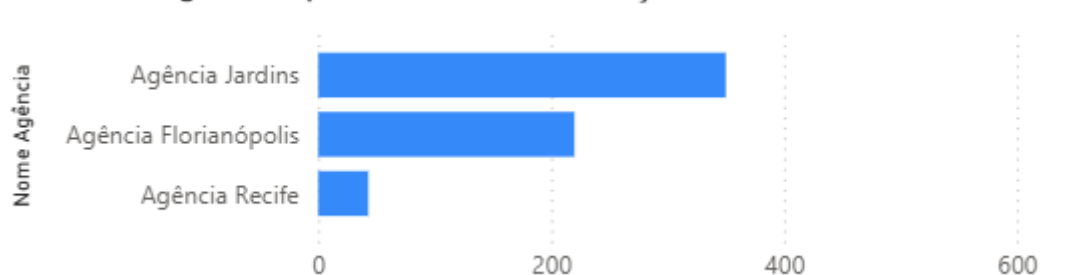
Item a – 5

A análise de desempenho das agências dos últimos 6 meses mostrou grande disparidade no volume de transações, confirmando suspeita de distribuição desigual, mesmo desconsiderando a agência digital. As Top 3 agências em performance foram a Agência Matriz, Porto Alegre e Tatuapé, e as que performaram no bottom 3 foram a Agência Jardins, Florianópolis e Recife com o menor volume de transação.

Top 3 Agências por volume de transações



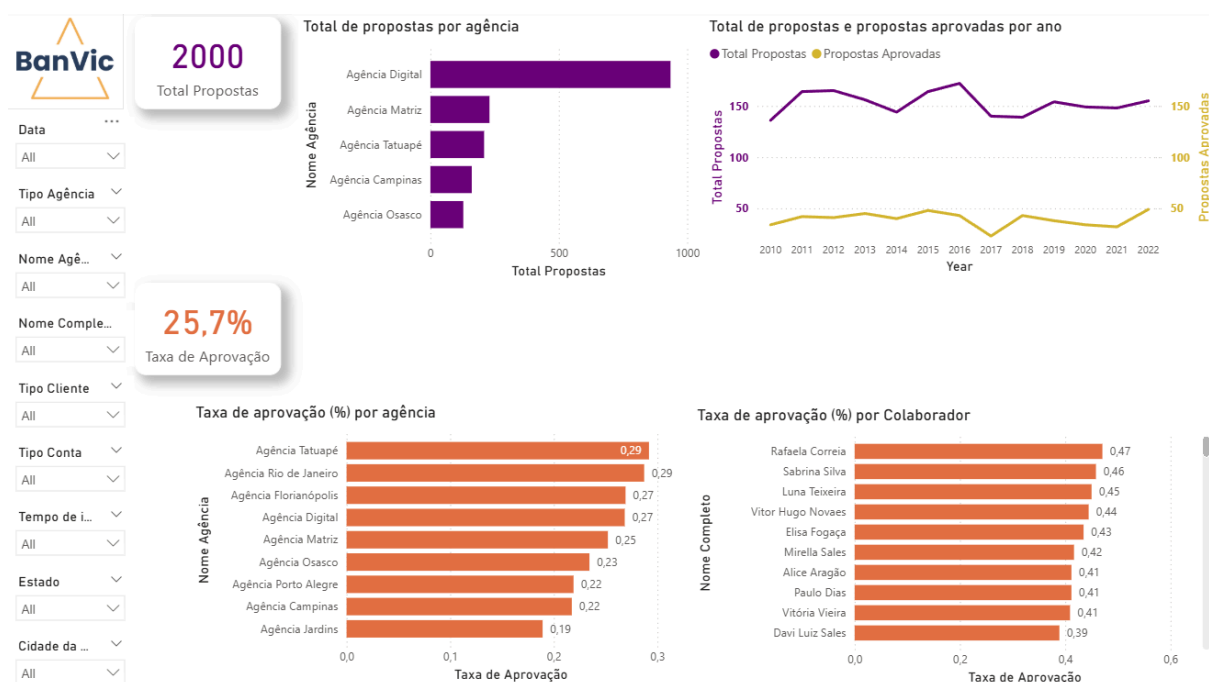
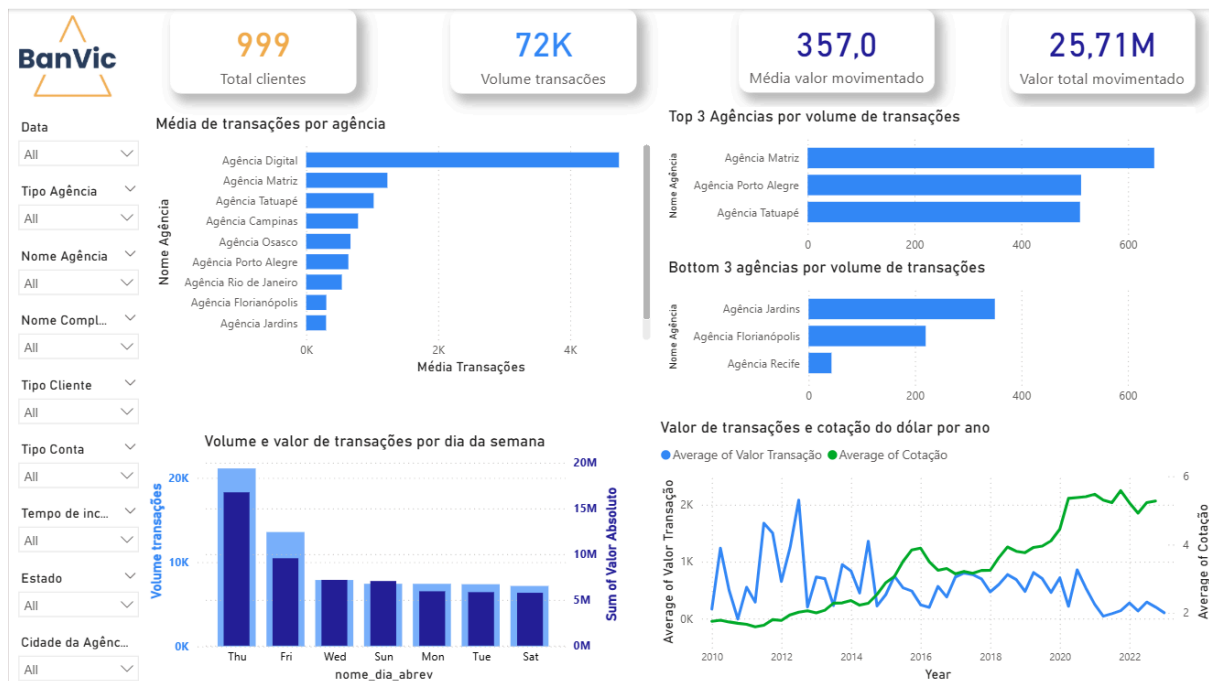
Bottom 3 agências por volume de transações



Itens C e D

Foram feitas análises exploratórias adicionais para conhecimento dos bancos de dados disponíveis utilizando Python e os notebooks estão no repositório do GitHub, bem como o arquivo .pbix do Dashboard criado com todas as análises e visualizações para esse desafio.

Link para o repositório: https://github.com/taisthomsen/desafio_lh_indicium



6. Recomendações e Encaminhamentos

O Banco Victória S.A. enfrenta o desafio de aprimorar sua infraestrutura de dados para garantir maior confiabilidade, consistência e capacidade de análise. Atualmente, limitações na qualidade e integração dos dados dificultam a identificação de oportunidades estratégicas, especialmente no que se refere ao desempenho das agências e à influência de fatores externos.

Uma infraestrutura moderna de dados permite simular cenários (ex.: impacto de campanhas em canais digitais vs. físicos) e suportar decisões baseadas em evidências. Portanto, recomenda-se:

1. Dados e Governança

- Investir em melhoria da qualidade de dados (pipeline e tratamento de nulos).
- Criar processos de monitoramento contínuo (testes automáticos de qualidade via dbt).

2. Estratégia Comercial

- Redirecionar parte do investimento em marketing para fortalecer a agência digital, que concentra maior valor transacionado.
- Avaliar o papel das agências físicas em cidades menores → foco em relacionamento e crédito.

3. Fatores Externos

- Incorporar dados macroeconômicos (câmbio, Selic, feriados) para modelos preditivos.
- Investir em análises estatísticas avançadas para validar correlações observadas.

4. Próximos Passos

- Consolidar o Data Warehouse como fonte única de verdade (*single source of truth*).
- Expandir integrações externas e avançar para análises preditivas (forecasting de demanda, propensão a crédito).