

## **Desafio Lighthouse - Análise de dados - Relatório final – Renan Pan**

Prezado(a) responsável pela avaliação, o seguinte relatório segue a estrutura fornecida juntamente ao arquivo do desafio, sendo os resultados então apresentados na seguinte ordem:

- a) As respostas do item 1, 2 e 3.
- b) Um tópico descrevendo brevemente quais os processos de transformação e tratamento de dados foram aplicados por você.
- c) Um dashboard dos principais KPIs de negócio do BanVic permitindo um detalhamento por data, por agência e por clientes.
- d) Apresentar as análises que foram necessárias para o entendimento do negócio e que podem ser úteis no suporte para a tomada de decisão.
- e) E finalmente, com base nas suas análises, uma conclusão com recomendações/sugestões de negócio e encaminhamentos para a CEO do BanVic.
- f) Apresentar e justificar as ferramentas adotadas para a elaboração das análises.

### **1. As respostas do item 1, 2 e 3.**

1 - Com base na análise exploratória realizada nos dados apresentados, quais perguntas de negócio podemos responder para mostrar o valor que dados possuem para a empresa?

- Alguns pontos podem ser avaliados como:
  - Qual o perfil do cliente?
  - Como está o desempenho das agências?
  - Transações realizadas (quantidade e tipo)
  - Colaboradores
  - Tendências temporais
  - Eficiência operacional

2 - Quais serão os indicadores ou visualizações utilizadas para responder às perguntas de negócio citadas anteriormente?

- Perfil do cliente
  - Visualização de mapa da localização e total de clientes por estado;
  - Idade média e distribuição da idade dos clientes.
  - Detalhe para média, mediana e desvio padrão do saldo total dos clientes.
- Desempenho das agências
  - Gráficos para a exibição do saldo médio por agência
  - Gráfico para exibição do número de contas por tipo de agência.
- Colaboradores
  - Gráfico de mapa da localização e total de colaboradores por estado.
  - Gráfico de dispersão com o total de clientes por colaborador
  - Gráfico de dispersão com o saldo total trabalhado por colaborador.
- Análise de transações
  - Gráfico de pizza para mostrar os tipos de transações para o período total

- Gráfico de linha com o total de transações e o saldo ao longo do tempo.
- Tendências temporais
  - Gráfico de área para as transações realizadas ao longo dos anos
  - Gráfico de área para os status de financiamentos ao longo dos anos
  - Gráfico de caixa para distribuição da taxa de juros (ao mês) média ao longo dos anos.
- Eficiência operacional
  - Análise de coorte para total de gasto com transações
  - Análise de vida útil da coorte
  - Análise de coorte para retenção de clientes para transações.

3 - Apresentar pelo menos duas análises de negócio para cada uma das categorias. Pergunta de nível teórico.

- Uma análise descritiva;
  - Calcular média, mediana e desvio padrão do saldo total por tipo de conta.
  - Gerar um gráfico de barra para mostrar os tipos de agências
- Uma análise diagnóstica;
  - Identificar padrões temporais nas transações diretas dos clientes
  - Identificar padrões nos valores totais de financiamentos ao longo dos anos
- Uma análise prescritiva;
  - Implementar um modelo preditivo para prever o impacto de alterações operacionais no saldo disponível.
  - Utilizar modelos de séries temporais para prever tendências futuras nas transações.
- Uma análise preditiva.
  - Utilizar análise de séries temporais para prever a demanda futura por propostas de financiamento.
  - Desenvolver modelos de regressão para prever a probabilidade de aprovação de propostas com base em diferentes variáveis

Na sequência trago a apresentação dos gráficos e análises realizadas com base no conjunto de dados fornecido.

## **2. Um tópico descrevendo brevemente quais os processos de transformação e tratamento de dados foram aplicados por você.**

A extração e tratamento dos dados foi realizada tanto no Python quanto no Tableau, sendo avaliados caso a caso os campos de informação e convertidos para o modelo apropriado, seja ele de data, hora ou algum outro formato necessário. Pelo Python com a biblioteca Pandas consigo anexar os dados e através do método .info() consigo ver qual o tipo dos dados do DataFrame, alterando-os se necessário. Pelo Tableau, ao carregar os arquivos

csv, conferi coluna por coluna se os dados se encontravam no formato certo e editei os mesmos quando foi necessário.

### **3. Um dashboard dos principais KPIs de negócio do BanVic permitindo um detalhamento por data, por agência e por clientes.**

Os dados estão distribuídos em 3 dashboards para visualização no Tableau Public. Através do campo superior direito é possível filtrar por agência (nome da agência) e data.

1º Dashboard – Colaboradores e clientes por agência e estado ao longo dos anos

- [https://public.tableau.com/shared/7P64WMW2H?:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/shared/7P64WMW2H?:display_count=n&:origin=viz_share_link)

2º Dashboard – Transações realizadas por ano e saldo total das agências conforme o ano

- [https://public.tableau.com/shared/B6SCD4CZC?:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/shared/B6SCD4CZC?:display_count=n&:origin=viz_share_link)

3º Dashboard – Valores disponibilizados para financiamentos ao longo dos anos; taxa de juros e número de propostas conforme saldo.

- [https://public.tableau.com/views/DA\\_challenge/Dashboard\\_financiamentos?:language=pt-BR&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/DA_challenge/Dashboard_financiamentos?:language=pt-BR&:display_count=n&:origin=viz_share_link)

Demais informações referentes ao projeto podem ser encontradas no repositório github: [https://github.com/renanpan/lighthouse\\_DA](https://github.com/renanpan/lighthouse_DA)

### **4. Apresentar as análises que foram necessárias para o entendimento do negócio e que podem ser úteis no suporte para a tomada de decisão.**

A partir deste ponto entram as análises realizadas tanto no Python como no Tableau para entendimento do negócio e para a tomada de decisão.

#### **4.1 Qual o perfil do cliente?**

O banco BanVic possui, desde 2010, 998 clientes localizados em sua grande maioria no estado de São Paulo (Figura 1).

### Cientes por estado

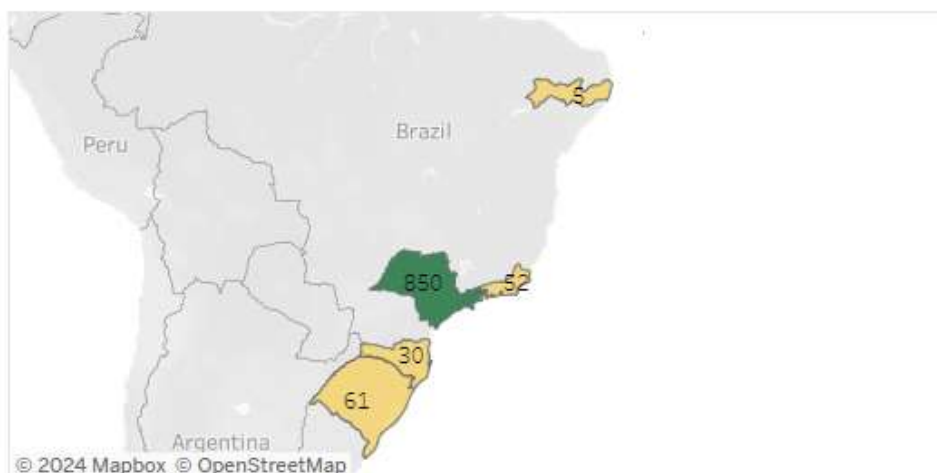


Figura 1 - Total de clientes do BanVic por região no Brasil.

Considerando-se a data de 05/02/2024 como base para cálculo, a idade média dos clientes do banco BanVic é de 49 anos. Com base na figura 2 pode-se observar que as idades dos clientes se encontram entre 16 e 81 anos.

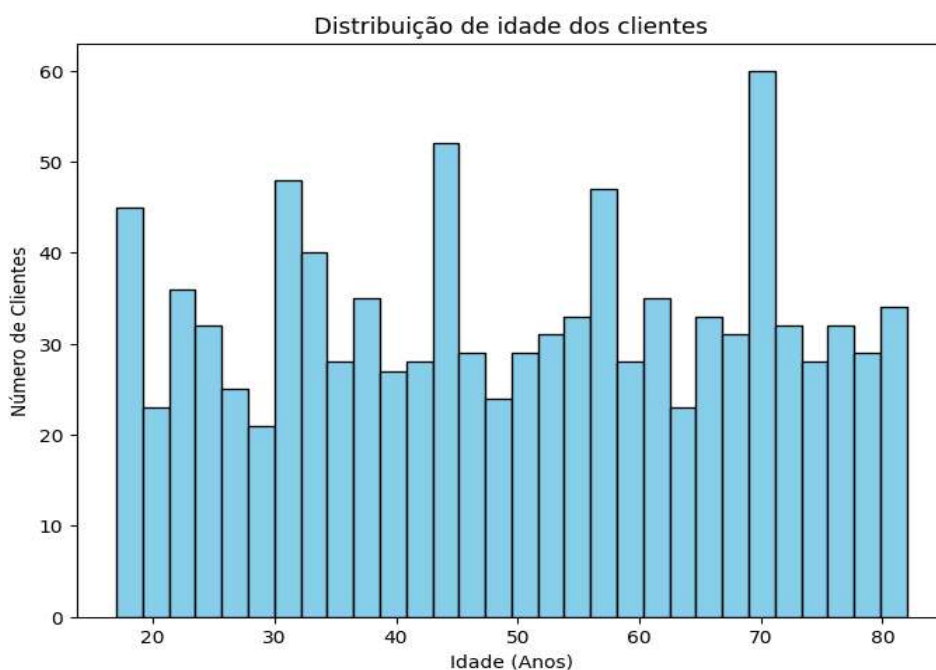


Figura 2 - Distribuição das idades dos clientes do BanVic

A figura 3 nos traz a distribuição da aderência de clientes ao BanVic desde 2010. Com base nos dados é possível observar que há um crescimento linear no total de clientes do banco e que, desde 2016 a maioria dos clientes vem através de agências digitais.

Clientes por agência (física e digital) por ano



Figura 3 - distribuição da adesão de clientes ao longo dos anos conforme a agência.

#### 4.2 Desempenho das agências

É possível averiguar que a agência digital é a que possui o maior número de clientes (Figura 4), assim como o maior saldo (Figura 5). Entretanto, ao avaliar o saldo médio por agência (Figura 6), temos as agências Matriz e Tatuapé como com maior saldo médio. Além disso, após a agência digital, a agência Matriz e Tatuapé são as que possuem o maior número de clientes.

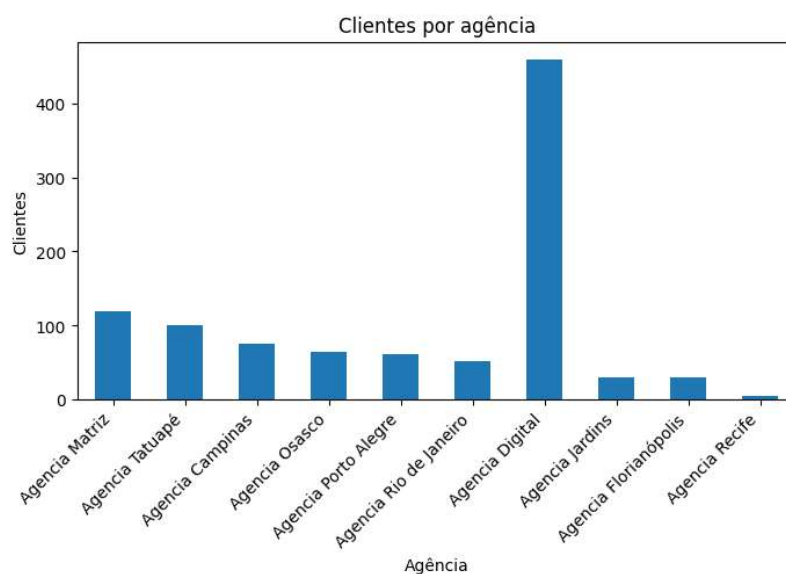


Figura 4 – Total de clientes por agência.

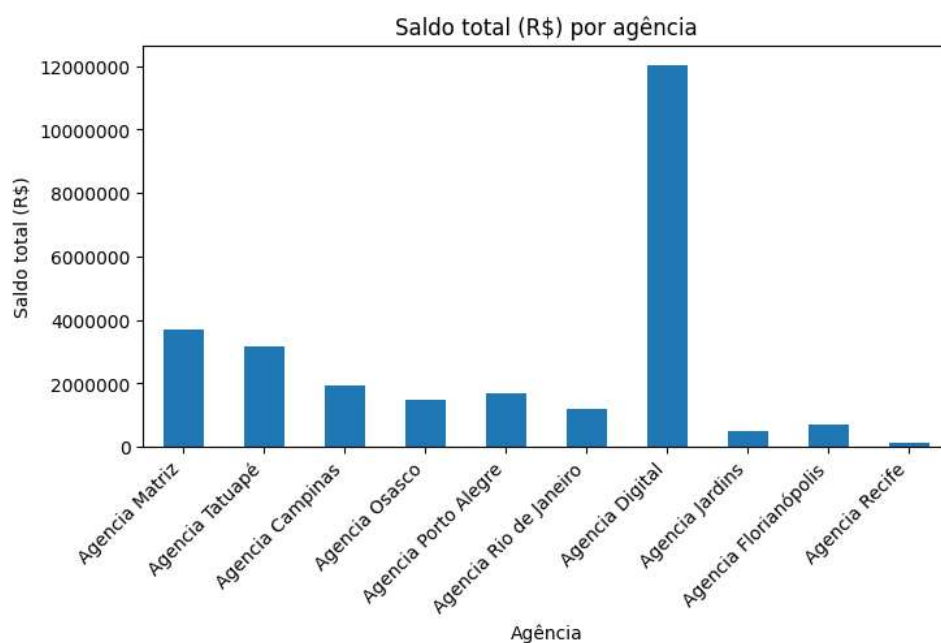


Figura 5 – Saldo total (R\$) por agência.

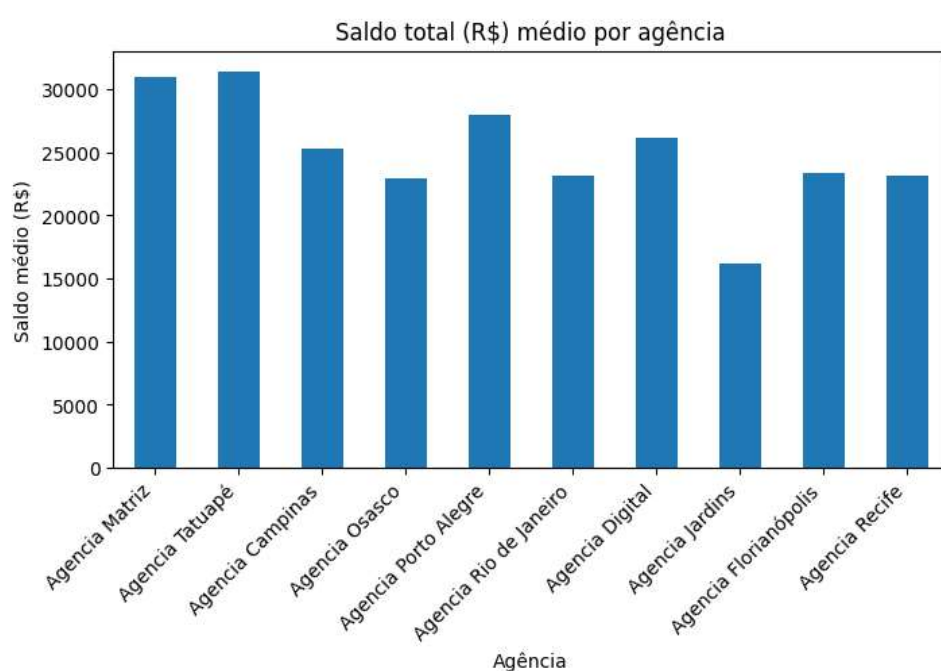


Figura 6 – Saldo total médio por agência

### 4.3 Colaboradores

As agências possuem um total de 100 colaboradores, os quais foram contratados entre os anos de 2010 e 2021 (Figura7). Os quais estão em sua maioria localizados no estado de São Paulo (Figura 8).

Colaboradores por agência por ano de data de abertura da agência

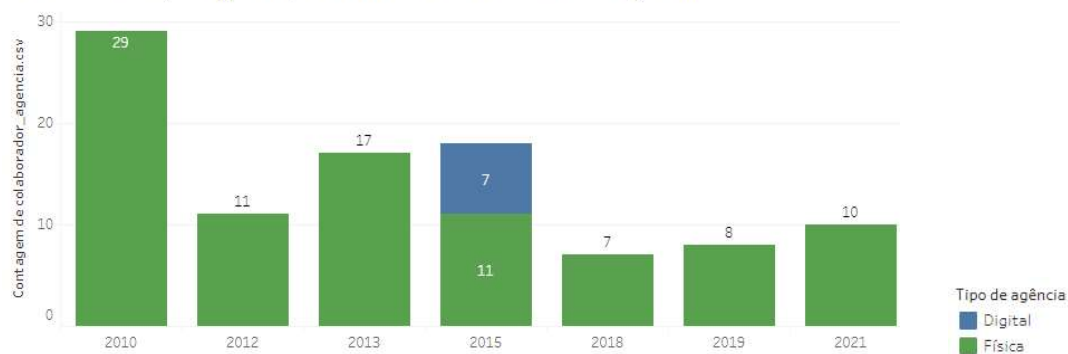


Figura 7 – Colaboradores por agência e por ano de abertura da agência.

Colaboradores por estado



Figura 8 - Distribuição dos colaboradores do banco BanVic por região do Brasil.

Quando consideramos o total de clientes que cada colaborador possui (Figura 9), podemos observar que sua grande maioria atende até 20 clientes, sendo que em alguns casos há colaboradores atendendo mais de 60 clientes.

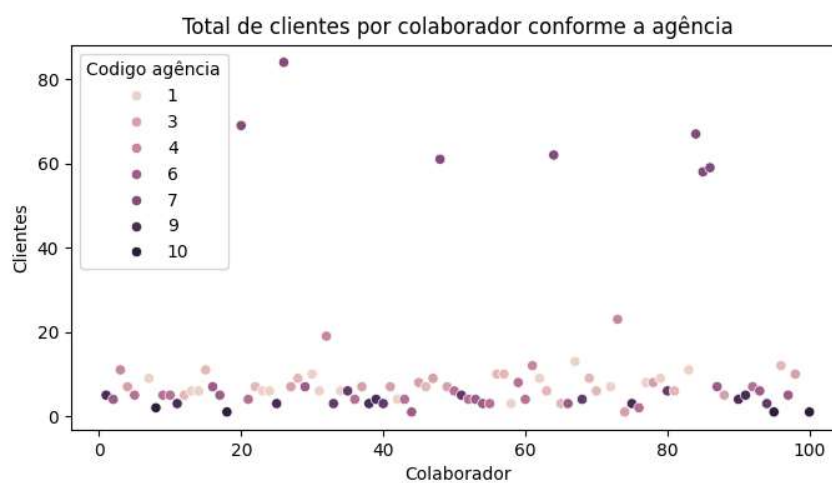


Figura 9 – Clientes por colaborador conforme a agência.

Além disso, ao analisarmos o saldo que cada um dos colaboradores trabalha (Figura 10), pode-se observar que a maioria dos colaboradores não trabalha com saldos acima de R\$ 500 mil.

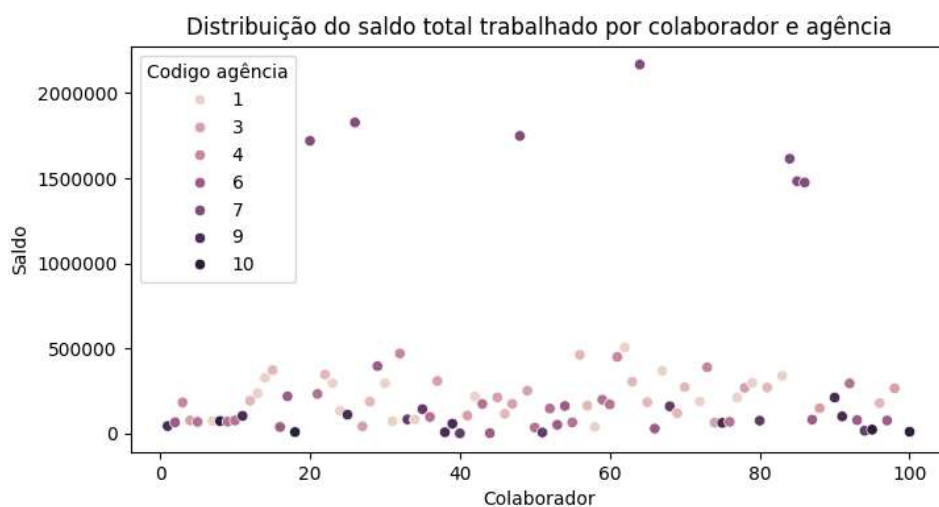


Figura 10 - Distribuição do saldo total trabalhado por colaborador e agência.

#### 4.4 Transações

Ao considerarmos todas as transações realizadas em todas as agências, temos que a maior quantidade de transações que ocorre é a compra no crédito (34,69%), seguido do débito (24,53%) e em terceiro a realização de Pix (18,28%). Demais transações envolvendo transferências de valores (que não incluem o Pix), estão em sua maioria abaixo 1,4% de participação nas transações (Figura 11).

Total de transações realizadas nas agências ao longo de todo o tempo

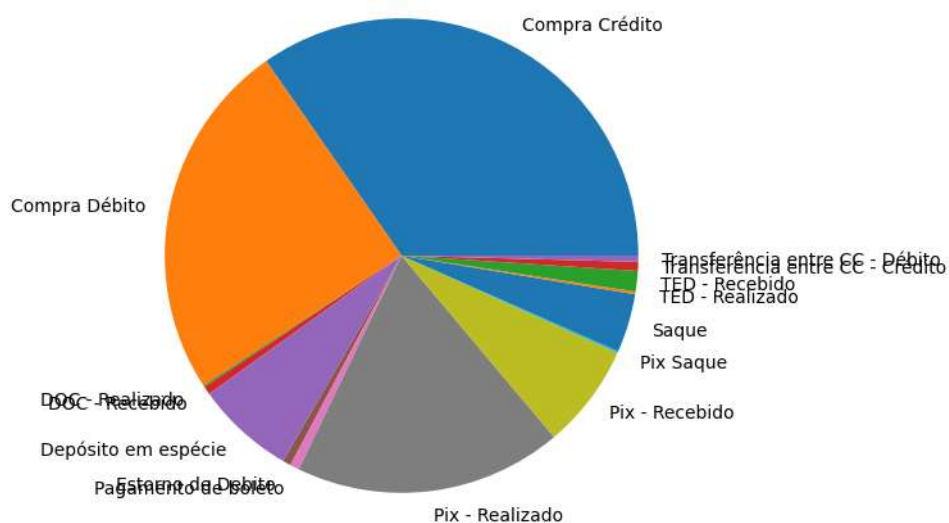


Figura 11 – Total de transações realizadas nas agências ao longo de todo o tempo.



A Figura 12 nos mostra que há uma evolução crescente nas transações realizadas no banco o que leva ao aumento no saldo total. Dentre as transações a compra no crédito e compra no débito são as que apresentam crescimento linear mais estável e, a partir de 2020 com o surgimento do PIX, podemos observar uma alta utilização deste meio por parte dos usuários.

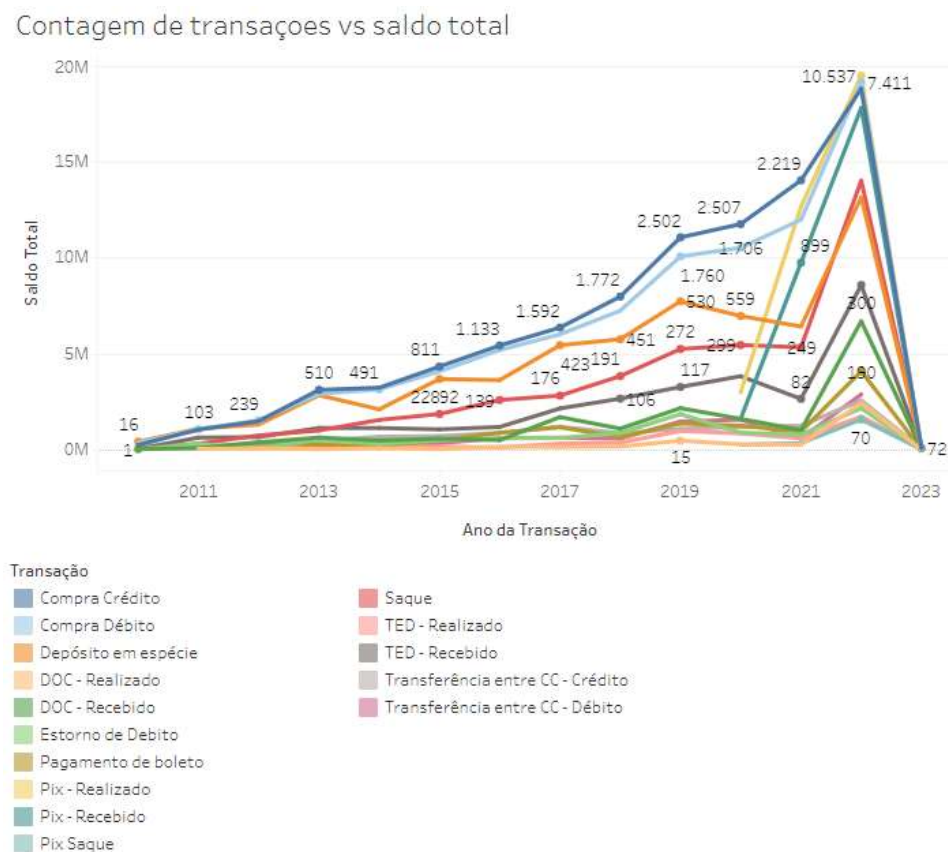
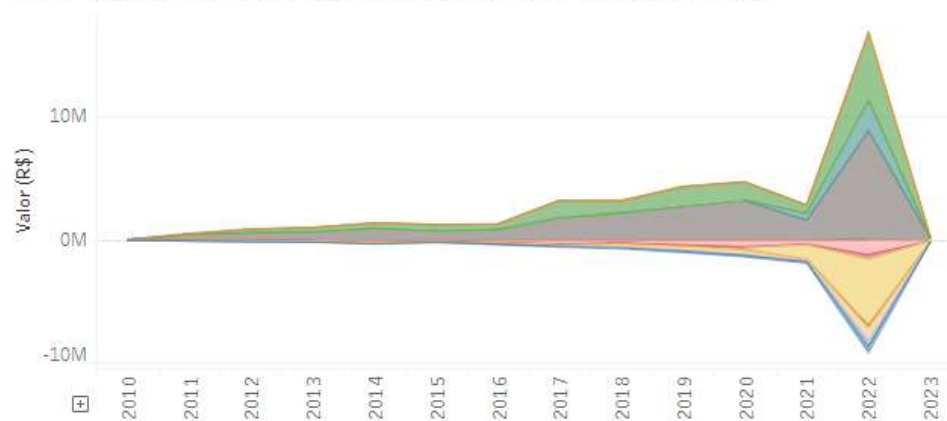


Figura 12 – Contagem de transações financeiras e saldo total do banco ao longo do tempo.

Um dos pontos que vale a pena ser destacado é que a implementação do PIX levou em 2022 a uma alta variação no total de entrada e saída de capital do Banvic, onde o pix realizado foi a maior fonte de saída de dinheiro do banco, enquanto TED e DOC recebidos a maior fonte de entrada de dinheiro no banco, contribuindo assim para a obtenção de um saldo positivo (Figura 13).

Valor gasto em transações financeiras conforme o ano



Contagem de transações por ano

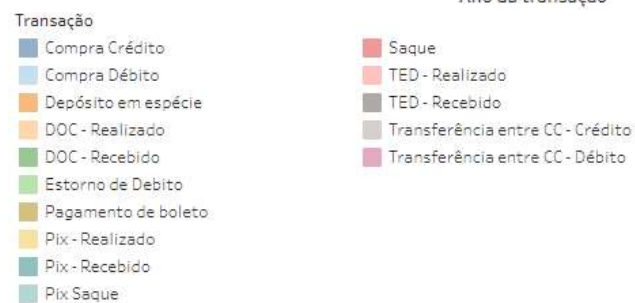
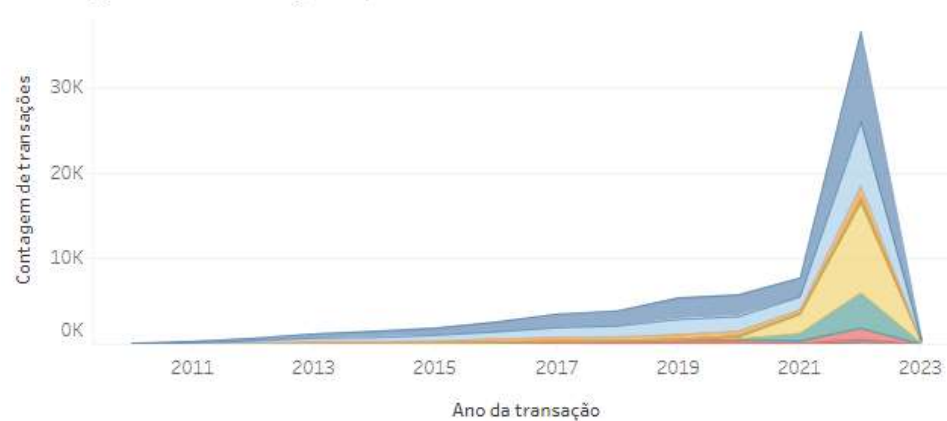


Figura 13 – Valor gasto em transações financeiras por ano e contagem de transações financeiras por ano.

## 4.5 Financiamentos

Desde sua fundação, o BanVic teve um total de 2000 propostas de financiamento, distribuídas no conjunto de dados em: ‘Enviada’, ‘Aprovada’, ‘Validação de documentos’ e ‘Em análise’ (Figura 14).

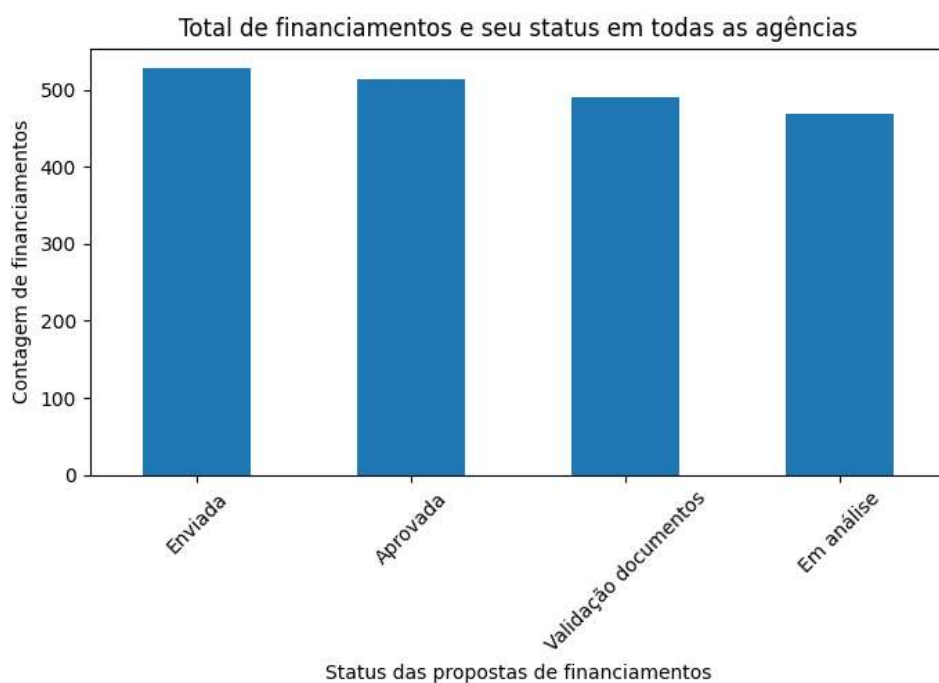


Figura 14 - Distribuição do status das propostas de financiamento do BanVic desde 2010.

Cada um dos status de financiamento ocupou cerca de R\$ 40 milhões do caixa do banco desde 2010 (Figura 15).

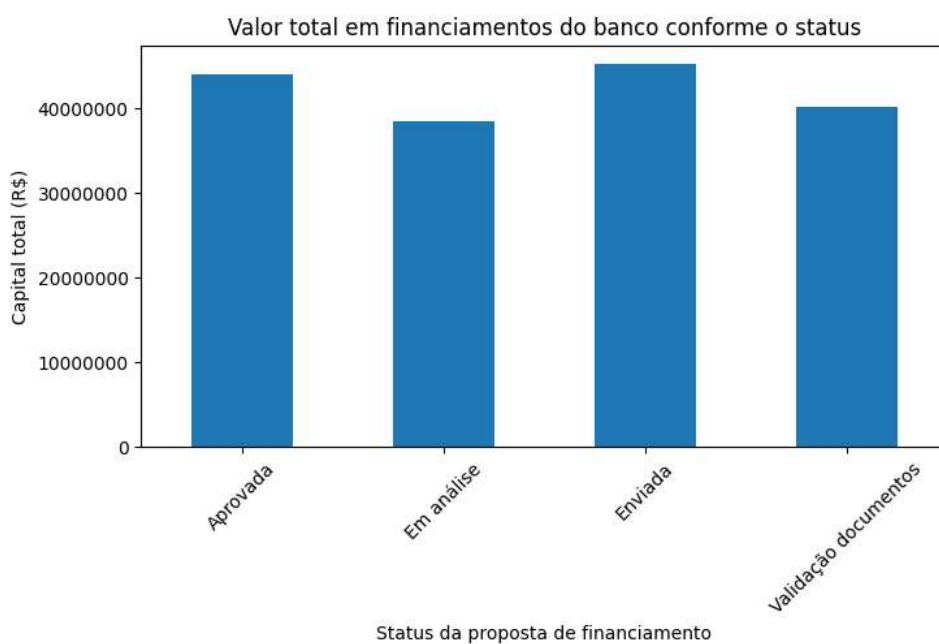


Figura 15 – Valor total em financiamentos do banco conforme o status da proposta.

Ao separar os valores totais para financiamento ao longo dos anos, pode-se observar que em cada ano o banco precisa de no mínimo R\$ 11 milhões para financiamentos (Figura 16).

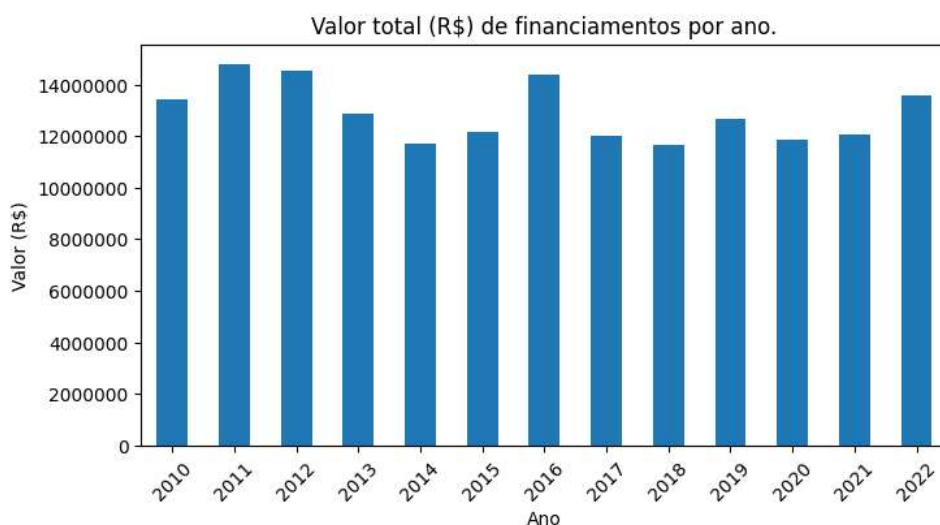


Figura 16 – Valor total (R\$) em financiamentos por ano.

Ao analisarmos a distribuição dos financiamentos (Figura 17) podemos ver que a maioria dos financiamentos pedidos no banco se encontra na faixa entre R\$50.000,00 e R\$110.000,00. Com alguns casos acima de R\$ 175.000,00. Isso corrobora com os dados da figura 18, que mostram uma constância nos valores do banco separados para cada um dos status dos financiamentos, mostrando que o banco trabalha de forma mais segura.

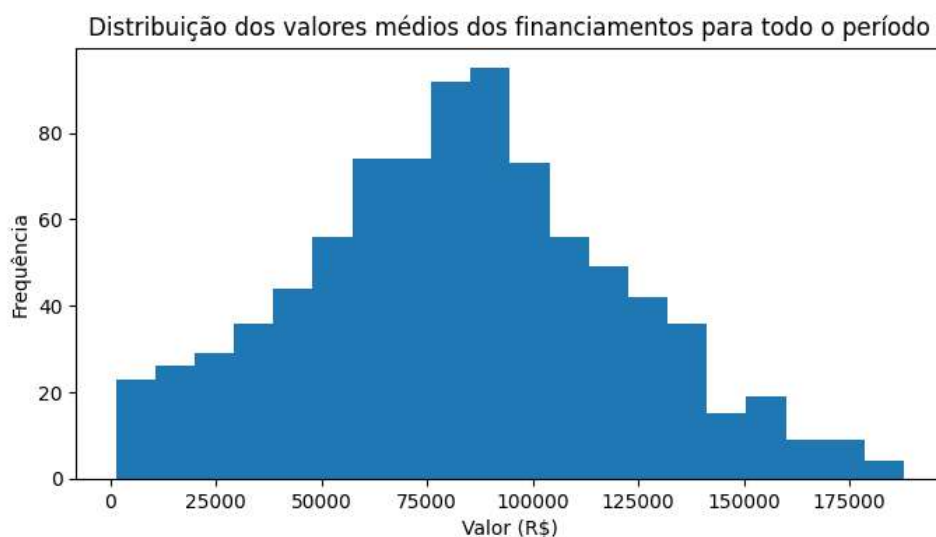


Figura 17 - Distribuição dos valores médios dos financiamentos para todo o período.

Valores médios de entrada e financiamento conforme o ano.

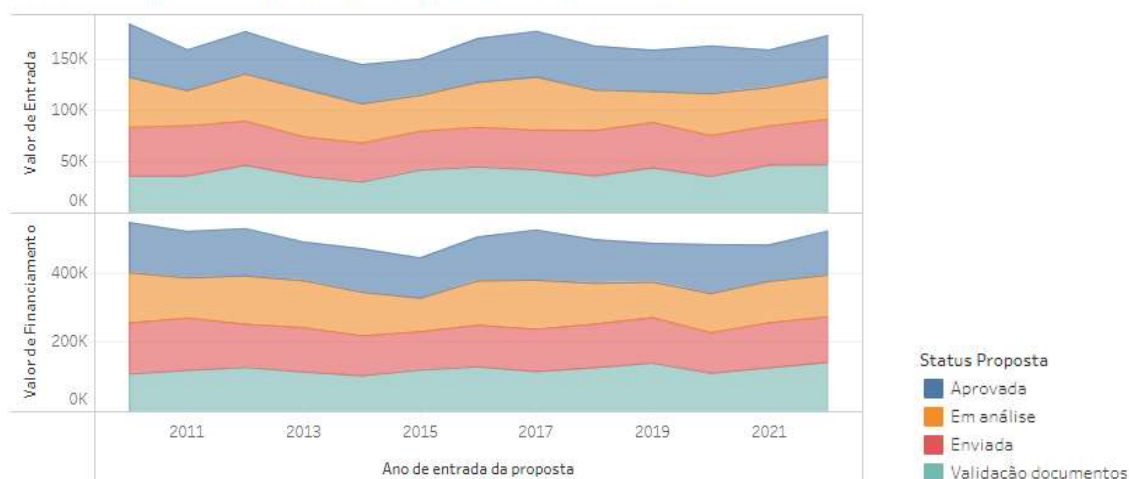


Figura 18 – Valores médios de entrada e financiamentos conforme o ano da proposta.

Outro fator que mostra a segurança do banco em trabalhar com é a quantidade de propostas de financiamento versus o saldo total onde o banco também mostra sua eficiência, tendo em sua maioria mais propostas aprovadas do que em outras fases (Figura 19).

Propostas de financiamento vs saldo total



Figura 19 - Propostas de financiamento versus saldo total ao longo dos anos.

No geral, as taxas de juros ao mês para os financiamentos do banco para todo o período desde 2010 mostram pouca oscilação, ficando em torno de 1,5% e 1,9% (Figura 20).

Valores médios para Taxa de juros (mensal) dos financiamentos

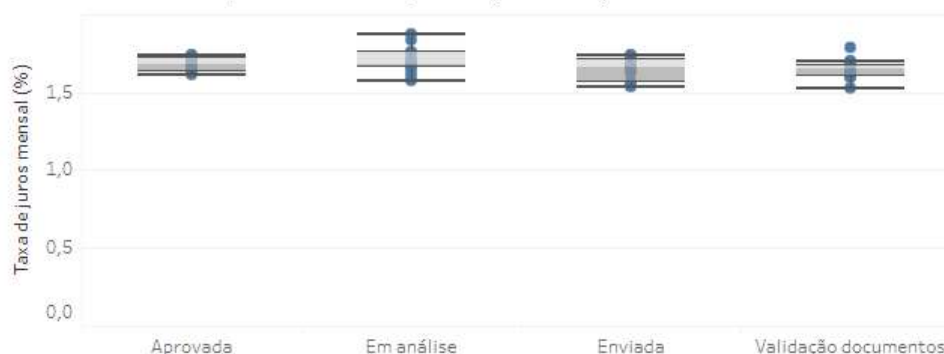


Figura 20 – Valores médios para taxa de juros (mensal) dos financiamentos.

## 4.6 Diferenças entre agência física e digital

Para comparação entre as transações e financiamentos das agências físicas e digitais, foram utilizados os dados do ano de 2022 em função da quantidade de dados e para incluir transações envolvendo PIX.

Quando consideramos as principais transações realizadas em agência física ou digital temos comportamentos semelhantes entre as transações realizadas pelos clientes. No ano de 2022 na agência digital (Figura 21) a principal transação realizada é a compra no crédito, enquanto na agência física (Figura 22) a principal transação realizada foi a realização de Pix.



Figura 21 – Total de transações realizadas nas agências digitais em 2022.



Figura 22 – Total de transações realizadas nas agências físicas em 2022.

Análise estatística entre as agências físicas e digitais.

- Existe diferença entre os saldos de clientes que utilizam as agências físicas e agências digitais?
  - o Foi realizado o teste t-student e obtido um p-value de 0,00322 e com base neste resultado rejeitou-se a hipótese nula de que não há diferença entre os saldos dos clientes que utilizam agência física ou digital. Assim, conclui que existe diferença entre os saldos dos clientes que utilizam agência física ou digital para realizar transações bancárias.
- Existe diferença nos valores dos financiamentos entre os clientes que utilizam agência física ou agência digital?
  - o Com base nos resultados do teste t foi obtido um p-value de 0,40557 ou seja, não foi rejeitada a hipótese H0 de que não há diferença entre os valores dos financiamentos entre clientes que utilizam agência física ou digital.

#### 4.7 Análise de Coortes

Para entendimento dos dados conforme as transações realizadas pelos clientes, foi realizada a análise de coortes primeiramente para o total gasto por clientes desde sua primeira transação (Figura 23). Em função da quantidade de informação, escolhi trabalhar com os dados desde o ano de 2018 (Figuras 24, 25 e 26) que possibilitam uma melhor visualização.

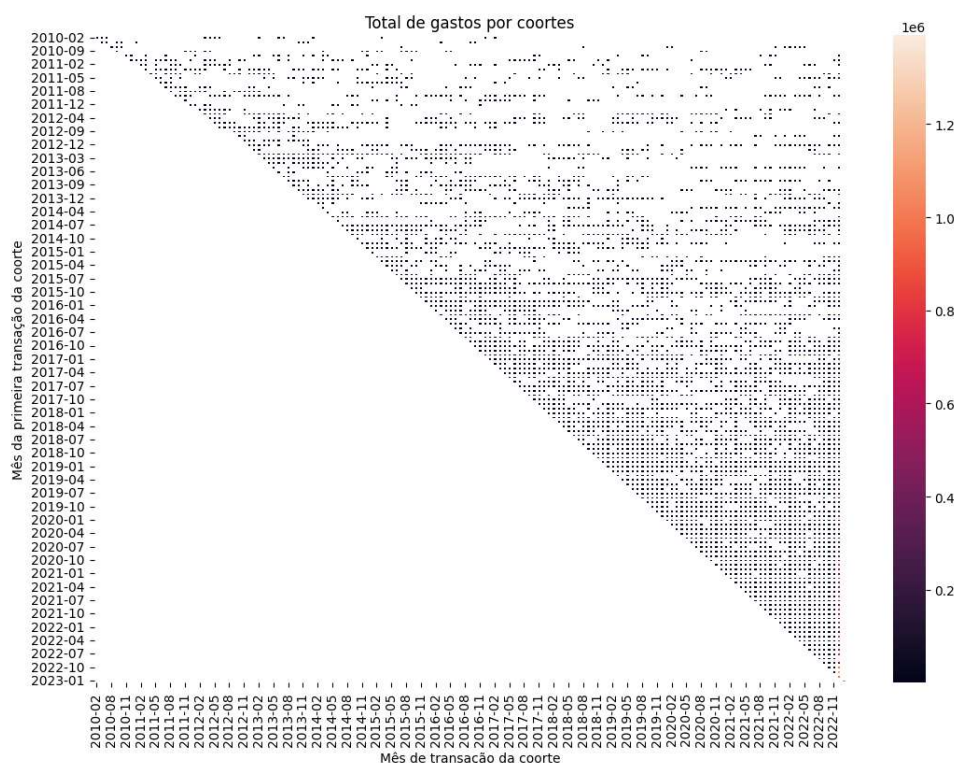


Figura 23 – Análise de coorte para o total de gastos pelos usuários desde 2010.

Com a filtragem dos dados para compreender os valores após 2018 temos coortes onde os gastos chegaram a cerca de R\$ 1,4 milhões em alguns casos e que o período entre novembro de 2022 e janeiro de 2023 foi o mês com maiores gastos realizados em transações para clientes das coortes a partir de maio de 2019 (cores mais claras no gráfico) (Figura 24).



Isso possivelmente ocorreu em função da época do ano ser mais próxima às festividades de fim de ano.

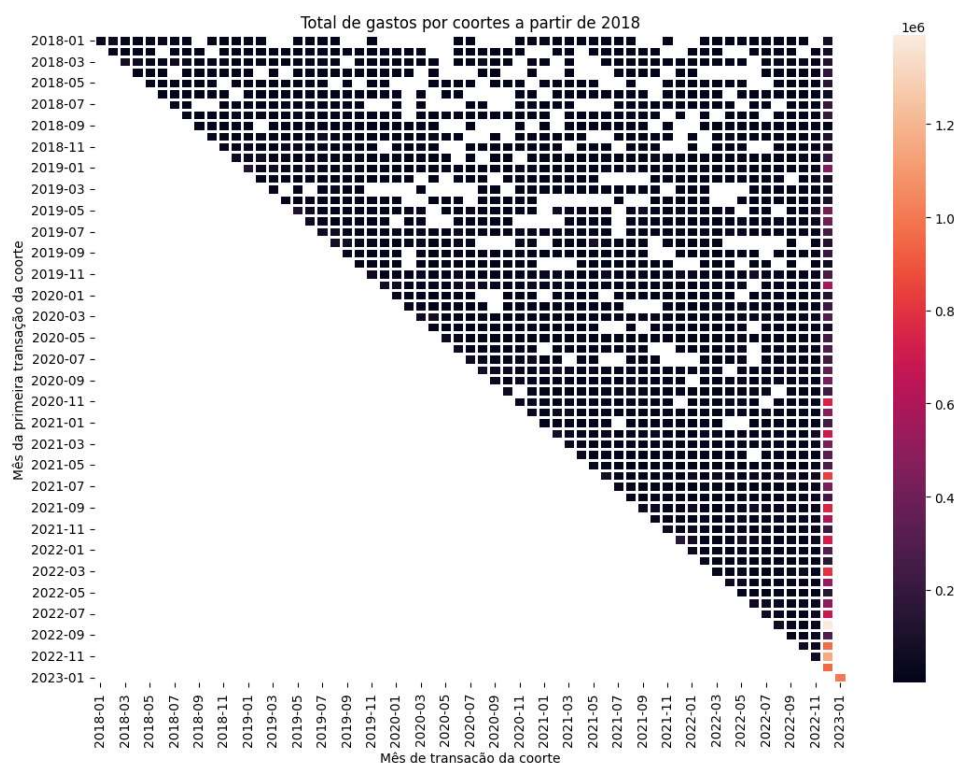


Figura 24 – Total de gastos por coortes a partir de 2018.

A figura 25 traz o tempo de vida das coortes. Com base nos dados podemos observar que em alguns pontos específicos os integrantes de cada coorte não realizaram nenhuma transação. Além disso, para todas as coortes, a maior quantidade de transações foi realizada próximo ao final do período de meses desde sua criação (cerca de 11/2022 e 01/2023), apresentando valores acima de 20 transações.

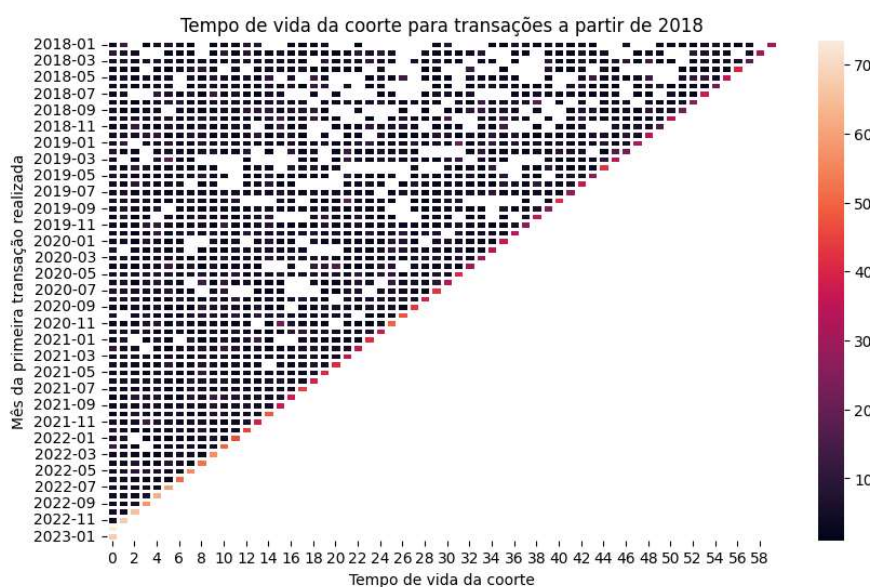


Figura 25 - Tempo de vida da coorte para transações a partir de 2018.



Outra informação interessante que podemos trazer com a análise de coortes é a taxa de retenção dos clientes com base nas transações realizadas. A figura 26 nos mostra que para as transações a partir de 2018 a taxa de retenção dos clientes está entre 0,4 e 0,6 (40 e 60%), sendo maior para clientes mais novos.

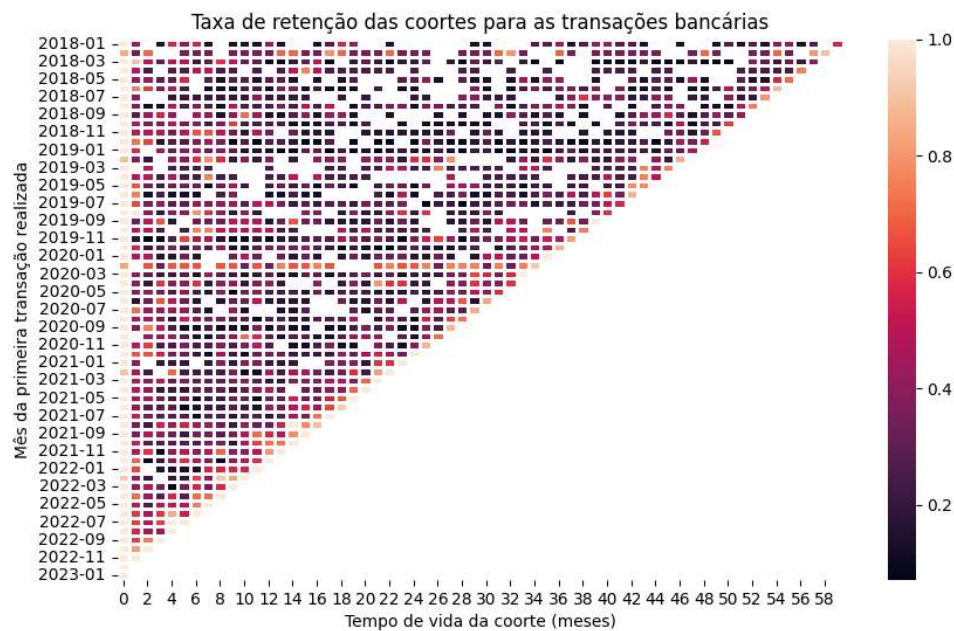


Figura 26 – Taxa de retenção das coortes para as transações bancárias

## **5. E finalmente, com base nas suas análises, uma conclusão com recomendações/sugestões de negócio e encaminhamentos para a CEO do BanVic.**

Os clientes do BanVic estão distribuídos na faixa entre 16 e 81 anos. A maioria deles se encontra no estado de São Paulo, mas tem clientes também no Rio Grande do Sul, Santa Catarina, Rio de Janeiro e Pernambuco. Aqui entra a primeira sugestão de que o banco poderia expandir sua área de atuação para outras capitais, como Curitiba uma vez que já possui clientes em São Paulo e Santa Catarina.

Agências com poucos clientes, como de Recife, apresentam um alto saldo médio mas alguma ação estratégica pode ser realizada a fim de melhorar a participação do BanVic na região.

Outra questão a ser observada é a quantidade de colaboradores na agência do Recife para a quantidade de clientes. São dez colaboradores e 5 clientes, a empresa tem potencial para atender mais pessoas.

Compras no crédito, débito e pix são as transações que mais vem crescendo ao longo dos anos. Como PIX e compras são duas formas de saída de capital do Banco, é possível que seja necessário um planejamento a longo prazo para manutenção do capital.

Com relação aos financiamentos, o banco vem se mostrando estável nos valores liberados para financiamento ao longo dos anos, sempre trabalhando dentro de uma margem de saldo total que o banco tem. Isso mostra que o banco vem seguindo seguro nas suas ações ao longo dos anos.

Ao conferir diferenças entre as transações de usuários de contas físicas e usuários de contas digitais, houve uma diferença no saldo das contas para cada tipo de conta, onde usuários digitais possuem maior saldo.

A análise de coortes mostra que a grande maioria dos clientes não chegou a realizar grandes compras, ficando em torno de até R\$ 200.000,00 gastos. Não houve algo que estimulasse os clientes a comprarem mais até o período entre novembro de 2022 e janeiro de 2023, onde os gastos chegaram próximos a R\$ 1,2 milhões. Nenhum outro período anterior a esse mereceu destaque. Isso provavelmente ocorreu em função do aumento crescente no número de clientes do banco com o passar dos anos. Se os dados assim continuarem espera-se maiores valores nos próximos anos.

Com relação a retenção de clientes com base nas transações desde 2018, temos que na maioria das coortes a taxa de retenção dos clientes está entre 40 e 60%, assim alguma estratégia pode ser realizada para angariar mais clientes.

## **6. Apresentar e justificar as ferramentas adotadas para a elaboração das análises.**

Para a análise estatística e exploratória dos dados utilizei o Python que me permite também a plotagem dos gráficos. Dentro dele utilizei algumas bibliotecas para gestão dos dados com Pandas e Numpy, e outras bibliotecas para trabalhar com gráficos como Seaborn,

Plotly.Express, Pyplot e também Scipy.stats para análise estatística. O Python também me serviu para fazer a análise de coortes, importante para entender o comportamento dos usuários com o tempo.

Para elaboração dos dashboards utilizei o Tableau Public que me permite fazer alguns cálculos e edição dos gráficos, além de permitir a filtragem dos dados, no caso por ano e data, para melhor entendimento de alguns períodos específicos.