

Desafio Análise de Dados /Engenharia de Analytics

NATÁLIA SILVA

natalia3.silva@alumni.usp.br

Entregas

Respostas do item 1, 2, 3 e 4.

1. Nem todos os envolvidos estão convictos com a ideia de uma frente de BIs no BanVic. Como podemos convencer Camila Diniz, diretora comercial, que as nossas soluções serão úteis para a empresa?

Dado o cenário descrito, onde diferentes personagens da empresa têm perspectivas variadas sobre o uso da análise de dados, é importante desenvolver uma estratégia que possa alinhar as visões e demonstrar o valor da análise de dados para todos os stakeholders, no intuito de ajudar a amadurecer a cultura e o uso de dados dentro da organização. À diretora comercial, Camila, em particular, esses dados têm grande valia, uma vez que ajudam na tomada de decisão e estabelecimento de novas diretrizes comerciais.

Abaixo, seguem pontos baseados na análise básica dos dados fornecidos, que podem instigar Camila a reconhecer a importância dessas análises:

- 1) Em específico para a área de marketing, Camila pode entender como as campanhas impactam a aquisição e retenção de clientes, permitindo medir o retorno sobre o investimento (ROI) e ajustar as estratégias para maximizar o alcance.

Considere um investimento hipotético de R\$ 100000 em marketing que foi efetuado nos últimos dois anos. Também levando em consideração a receita desse período, calculada através do valor das transações: R\$7708477.04 em 2022 e R\$57191.71 em 2023. Nesse cenário, temos um retorno (ROI) de 77%. Dessa forma, Camila pode justificar, através de dados, que os investimentos feitos em sua área derem retorno positivo ao BanVic.

- 2) Manter perto perfil de clientes que geram muita receita e direcionar campanhas de marketing para novos tipos de segmentos.

Com base nos dados fornecidos, identificamos que o tipo de cliente que gera mais receita é Pessoa Física, com um total de receita de R\$25698073.73. Assim, Camila pode buscar estratégias que mantenham PF perto e tragam também contas jurídicas para o BanVic.

- 3) Para a área comercial, também é importante entender a quantidade de clientes que se mantem ativos em um determinado período.

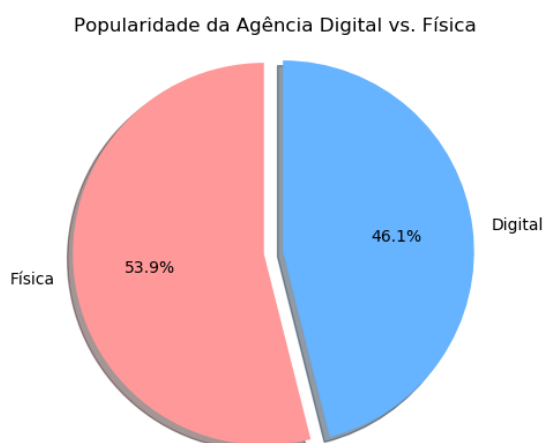
Os dados de ERP e CRM fornecidos permitem comparar o número de clientes que movimentaram sua conta em um período, comparado com um período anterior. Se considerarmos os anos de 2022 comparado com 2021, 89.28% dos clientes se mantiveram ativos. Esses dados podem direcionar as devidas equipes a tomarem medidas que retenham os clientes.

- 4) Entender o risco de churn e desenvolver estratégias de retenção baseado nos padrões de inatividade dos clientes e como isso afeta a receita.

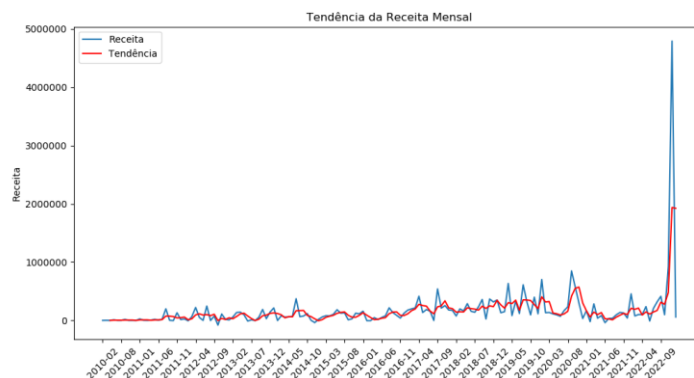
Por exemplo, através dos dados fornecidos, entendemos que 10.72% dos clientes deixaram de usar o serviço do banco entre 2021 e 2022.

- 5) Investir em um determinado tipo de agência mais lucrativo.

Sabemos que 54% dos clientes preferem agências físicas, contra 46% que preferem as digitais. Além disso, a receita das agências digitais e físicas são, respectivamente, R\$11579143.34 e R\$14126311.06. Portanto, a agência física gera 22.00% mais receita do que a agência digital.



- 6) Investigar variações e tendência na receita do BanVic que permita ajustes estratégicos. Na figura abaixo, é possível identificar a variação da receita média (baseada no valor das transações) e sua tendência/média móvel:



- 7) Avaliar a eficiência do processo de aprovação de crédito e identificar áreas para melhoria utilizando como base a taxa de conversão de propostas de crédito.

As planilhas enviadas mostram uma taxa de conversão de propostas de crédito de 25.70%. Assim, a equipe comercial pode identificar áreas para melhoria analisar os motivos de rejeição das propostas e outras métricas relevantes.

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

Para responder aos pontos levantados anteriormente, tomei por base os seguintes indicadores – na mesma ordem das questões levantadas:

1. *Retorno sobre o Investimento (ROI) em Análise de Dados*: Avalia o retorno financeiro em comparação com o investimento em tecnologias de análise de dados. Esse índice pode ser aplicado praticamente para todas as áreas de interesse. No exemplo da atividade anterior, nos baseamos em um investimento para a área de marketing.

Esse índice é calculado pela relação entre o retorno financeiro e o investimento, onde o “retorno” diz respeito à diferença entre a receita e o investimento em si.

2. *Customer Lifetime Value (CLV)*: Estima o valor total que um cliente traz para a empresa ao longo de seu relacionamento. É uma métrica que ajuda a identificar quais clientes são mais valiosos para a empresa e direcionar esforços de retenção e marketing.

No caso anterior, comparamos os tipos de cliente de Pessoa Física e Pessoa Jurídica.

3. *Taxa de Retenção de Clientes*: Mede a porcentagem de clientes que continuam ativos em um determinado período. Em especial, esse indicador pode

interessar à Camila Diniz (Marketing), que tem o objetivo de melhorar a segmentação dos clientes e aumentar o investimento em marketing.

Através de nossas tabelas de ERP e CRM, podemos identificar a quantidade de clientes que movimentou sua conta em determinado período (por exemplo, no ano de 2022), bem como num período anterior (ex 2021). Assim, obtém-se a taxa de retenção de clientes através da relação entre os clientes ativos no período em questão com relação ao anterior.

4. *Taxa de Churn (Cancelamento)*: Mede a porcentagem de clientes que deixaram de usar os serviços da empresa em um determinado período.

Filtrando os clientes que realizaram transações em um período estabelecido (ex ano de 2022) e a quantidade de clientes que o fizeram num período prévio (ex 2021), podemos calcular a taxa de Taxa de Churn através de sua diferença (clientes perdidos) sobre a quantidade de clientes ativos no primeiro período. É complementar ao índice anterior.

5. *Popularidade da Agência Digital vs. Física*: compara o uso da agência digital com a agência física. Uma simples relação entre os tipos de agência que estão informadas nas planilhas.

Esse indicador importa à Lucas Johnson (Análise Comportamental), por exemplo, pois tem por objetivo compreender o comportamento dos clientes em diversos canais.

6. *Crescimento da Receita*: Mede o aumento da receita decorrente de decisões informadas por análises de dados.

Calculamos a receita com base no valor das transações financeiras. Se tivermos interesse em receitas mensais, por exemplo, podemos calcular a porcentagem de mudança ao longo do tempo ou até mesmo sua tendência. Esse índice é de interesse, em particular, para Sofia Oliveira (CEO) que busca elevar o banco a novos patamares através da análise de dados.

7. *Tempo Médio de Aprovação de Crédito*: mede o tempo que leva desde a entrada de uma proposta de crédito até a sua aprovação ou rejeição. Este KPI é uma medida de eficiência do processo de aprovação de crédito dentro da empresa.

No nosso exemplo, não temos a data de aprovação das propostas e, portanto, calculamos apenas a taxa de aprovação, não o Tempo Médio de Aprovação de Crédito em si.

3. Uma dimensão de datas bem construída viabiliza a elaboração de análises mais robustas com relação ao tempo. O BanVic quer começar seu Data Warehouse (DW) com uma dim_dates que atenda seus requisitos de análise. Utilizando uma dimensão de datas, responda às duas perguntas a seguir. Por fim, proponha outras duas análises utilizando a dim_dates que possam ter valor de negócio para o BanVic.

- Qual dia da semana tem, em média, maior volume de transações e qual tem maior valor movimentado?
- O BanVic tem, em média, os maiores valores movimentados no início ou final de mês?

Com base na análise dos dados de transações do BanVic, identificamos que a quinta-feira é o dia com maior volume de transações (17978). Esse comportamento pode ser influenciado por diversos fatores como, por exemplo, a preparação para o final de semana ou mesmo a execução de transações comerciais que se acumulam ao longo da semana.

Por outro lado, quarta-feira é o dia da semana com o maior valor médio movimentado por transação (R\$ 398.82). Isso pode indicar que, nesse dia, os clientes do BanVic tendem a realizar relevantes decisões de investimento, grandes compras ou pagamentos corporativos.

Abaixo, segue a análise completa onde se é possível observar o valor médio, em real, e o volume das transações.

	weekday	volume_transacoes	valor_medio
0	Segunda-feira	4165	277.795971
1	Terça-feira	4380	91.029320
2	Quarta-feira	4784	398.818696
3	Quinta-feira	17978	204.437745
4	Sexta-feira	10538	200.950001
5	Sábado	4272	286.811547
6	Domingo	4508	367.829146

Já a análise mensal revelou que o início do mês é, em média, 10,26% mais movimentado do que no final do mês. Este resultado sugere que os clientes do BanVic tendem a movimentar mais valores nos primeiros 15 dias do mês em comparação com os últimos 15 dias, o que pode ser devido a fatores como recebimento de salários, pagamento de contas e despesas recorrentes, que geralmente ocorrem nesse período.

A utilização de uma dimensão de datas bem construída também nos permite fazer uma análise de sazonalidade, buscando entender como variam as transações ao longo do ano. Assim, o BanVic consegue antecipar períodos de alta e baixa demanda, permitindo uma melhor alocação de recursos e planejamento de campanhas de marketing sazonais. Abaixo, um exemplo da análise sazonal para o ano de 2022:



Adicionalmente, o BanVic também pode se beneficiar de uma análise de picos de atividade. Isto é, estudar se existem certos dias do mês que apresentam picos de atividade transacional. Identificar picos de atividade pode ajudar o BanVic a preparar melhor suas operações e sistemas para lidar com aumentos temporários na demanda, garantindo uma experiência do cliente mais suave.

4. André Tech, o especialista em tecnologia, solicitou que você buscasse fontes de dados públicas que poderiam ser utilizadas para enriquecer a base de dados do BanVic e ampliar as possibilidades de análise. Quais dados públicos você utilizaria e como você levaria esses dados para o Data Warehouse? Quais decisões poderiam ser facilitadas com o uso desses novos dados? Pergunta de nível teórico. O seu projeto final não necessariamente precisa trazer esses dados externos.

Para enriquecer a base de dados do BanVic e ampliar as possibilidades de análise, podem ser utilizadas várias fontes de dados públicos. Alguns exemplos de dados públicos que podem ser integrados ao DW:

- Dados Econômicos e Financeiros: Banco Central do Brasil, Instituto Brasileiro de Geografia e Estatística (IBGE), entre outros podem fornecer dados sobre taxas de juros, índices de inflação, PIB, taxa de câmbio ou indicadores de mercado financeiro.
- Dados Demográficos e Socioeconômicos: IBGE, DataSUS, ou outras fontes governamentais informam sobre população, renda média, nível de educação, índices de desenvolvimento humano (IDH).
- Dados Climáticos e Ambientais: de fontes como Instituto Nacional de Meteorologia (INMET), Instituto Nacional de Pesquisas Espaciais (INPE), ou

centros de pesquisa informam sobre temperaturas extremas, precipitação, eventos climáticos extremos.

- Dados de Saúde: provém do DataSUS ou Secretaria de Vigilância em Saúde (SVS) e informam sobre taxas de incidência de doenças, dados de hospitalização, campanhas de vacinação.

Para levar esses dados ao Data Warehouse do BanVic, podemos propor os seguintes passos:

- Identificação dos dados relevantes e seleção dos datasets que podem enriquecer a base de dados do BanVic.
- Extração dos Dados que pode ser feita através de APIs, downloads de arquivos CSV, XML, JSON, etc.
- Transformação e limpeza dos dados para garantir a consistência e integridade. É necessário converter formatos de data/hora, normalizar valores, remover duplicatas, tratar valores ausentes etc. Isto é, padronizar os datasets ao "template" proposto pelo BanVic.
- Integração dos dados no Data Warehouse, onde se faz o upload dos datasets seguindo a arquitetura ETL (Extract, Transform, Load).
- Atualizar a documentação e os dicionários de dados.

Adicionando esse conhecimento ao banco de dados da BanVic, é possível investigar em maiores detalhes externalidades que interferem nas transações bancárias. Por exemplo, ao utilizar taxas de juros, índices de inflação e outros indicadores, é possível melhorar os modelos de análise de risco de crédito e precificação de produtos financeiros. Além disso, torna mais assertivo a tomada de decisão sobre aprovação de crédito, definição de limites de crédito, taxas de financiamento.

Em termos de segmentação de clientes e marketing direcionado, os dados demográficos e socioeconômicos enriquecem a compreensão dos perfis dos clientes e auxiliam na criação de campanhas de marketing personalizadas, além de desenvolvimento de novos produtos e serviços.

Já questões ambientais contribuem para o planejamento e previsão de demanda, como ajustar estratégias de estoque e logística ou campanhas promocionais sazonais.

Monitorar tendências de saúde pública e ajustar produtos e serviços para atender às necessidades dos clientes contribui com o desenvolvimento de produtos de seguros de saúde, campanhas de bem-estar e saúde.

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

A princípio, é importante pontuar que as planilhas já vieram padronizadas e prontas para leitura. Assim, me poupou o trabalho de editar ou ler uma por vez.

Para minhas análises, utilizei Python e, em específico, os pacotes abertos pandas, numpy e matplotlib. Li e armazenei os arquivos CSV em DataFrames individuais para cada planilha – a minha tentativa de concatenar todas as planilhas em um dataframe não foi bem-sucedida.

As colunas de data foram convertidas para o formato datetime, que utilizei para realizar operações como calcular a diferença entre datas, tendências, etc.

Como armazenei as planilhas individualmente, precisei garantir que os nomes das colunas estavam equivalentes para comparar os dados transversalmente. E, a depender da análise que fiz, mergi dataframes com as colunas importantes. Além de mergir planilhas, também foi necessário agrupá-las pela variável de interesse como, por exemplo, tipo do cliente, tipo de agência, etc.

Nas análises que levavam em conta o número de transação ou número de clientes, precisei da função .unique() para garantir que não houvesse repetição.

Na atividade das dimensões de datas, incluí várias colunas derivadas da data, como ano, mês, dia, trimestre, semana do ano, dia da semana, indicador de fim de semana, dia do ano, indicador de feriado, indicador de dia útil e parte do mês. Os dados de transações foram convertidos para o formato datetime, e valores ausentes foram removidos para garantir a integridade dos dados. As colunas de data tanto nas transações quanto na dimensão de datas foram ajustadas para conter apenas a data (sem a hora) para garantir uma mesclagem correta.

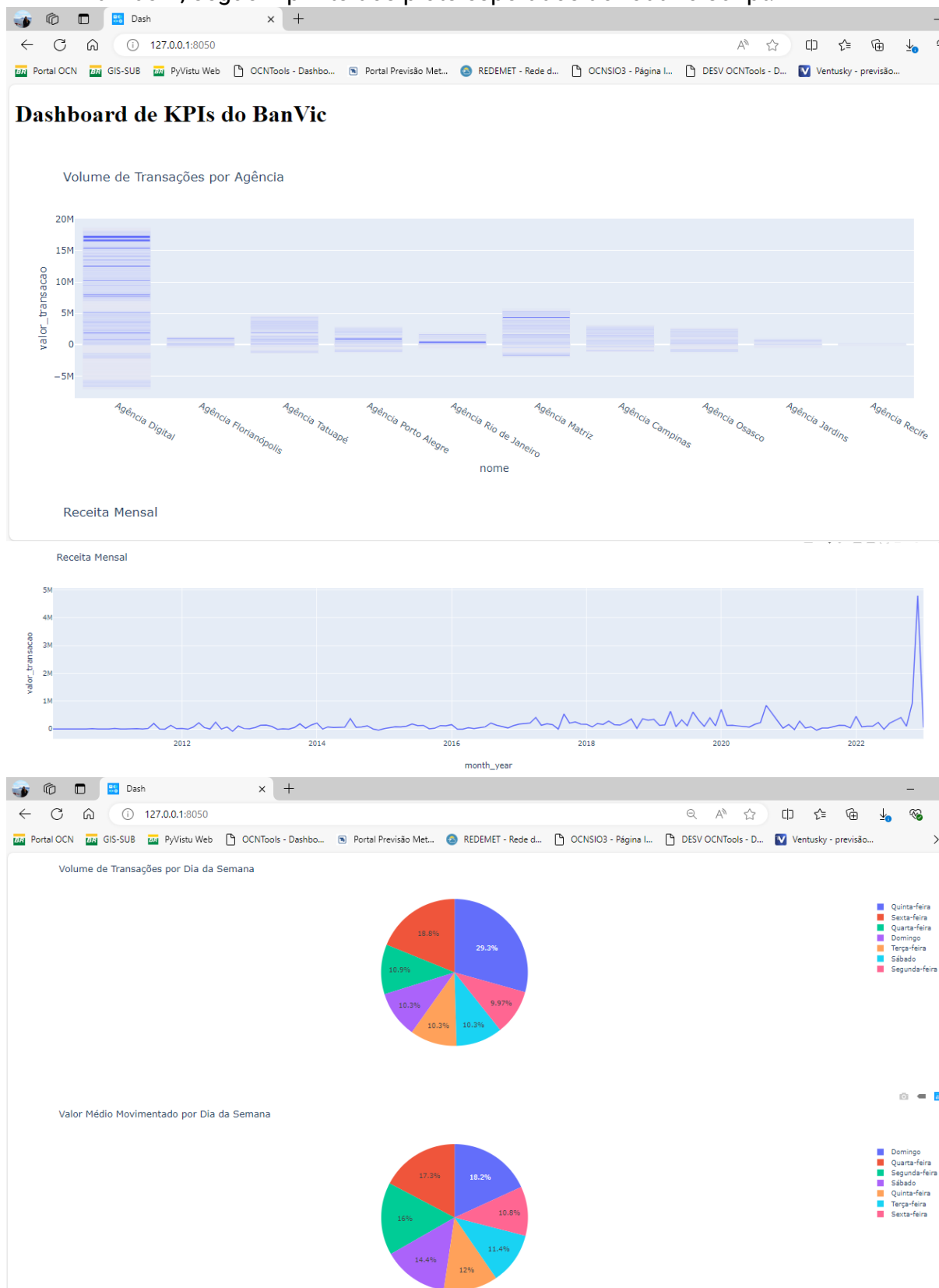
Foram realizados agregações e cálculos de métricas para responder às perguntas de negócio, como calcular a receita mensal, volume de transações e valor médio movimentado por dia da semana.

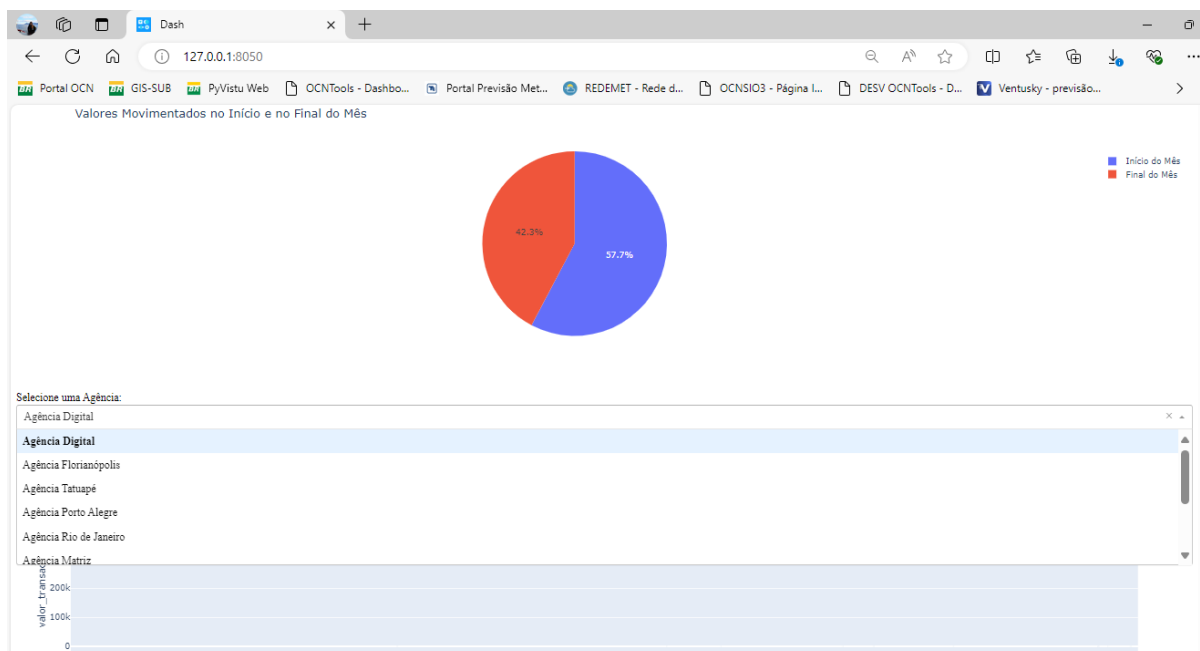
Um dashboard dos principais KPIs de negócio do BanVic permitindo um detalhamento por data, por agência e por clientes. Também apresente suas respostas às perguntas do item 3.

O dashboard foi construído dando seguimento nas análises do python e pode ser acessado rodando o código **dashboard.py** enviado em anexo (path: <http://127.0.0.1:8050/>; Debugger PIN: 018-846-560). Ao rodar o script, o terminal deve abrir as seguintes mensagens:

```
C:\Users\F300\OneDrive - Indiciu... X + -
Converting to PeriodArray/Index representation will drop timezone information.
Running on http://127.0.0.1:8050/
Debugger PIN: 793-689-955
* Serving Flask app "dashboard" (lazy loading)
* Environment: production
  WARNING: Do not use the development server in a production environment.
  Use a production WSGI server instead.
* Debug mode: on
cod_transacao num_conta data_transacao nome_transacao ... cidade uf data_abertura_y tipo_agen
cia
0 3109 256 2020-05-22 03:15:21+00:00 Saque ... São Paulo SP 2015-08-01 Digi
1 28485 256 2021-06-29 11:52:38+00:00 Saque ... São Paulo SP 2015-08-01 Digi
2 1776 512 2020-05-17 13:04:50+00:00 Saque ... Florianópolis SC 2019-10-09 Fis
3 14311 512 2020-09-06 02:55:34+00:00 Saque ... Florianópolis SC 2019-10-09 Fis
4 25012 512 2020-10-16 09:26:52+00:00 Saque ... Florianópolis SC 2019-10-09 Fis
[5 rows x 28 columns]
C:\Users\F300\OneDrive - Indiciu... \envs\ocnpy3\lib\site-packages\pandas\core\arrays\datetime.py:1172: UserWarning:
Converting to PeriodArray/Index representation will drop timezone information.
Running on http://127.0.0.1:8050/
Debugger PIN: 018-846-560
```


Também, seguem prints dos plots esperados ao rodar o script:





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

Diversas análises baseadas nas planilhas ERP e CRM foram realizadas para os dados do BanVic, de modo a entender melhor o negócio e fornecer suporte à tomada de decisão. Pontuo algumas das análises realizadas com resultados apresentados anteriormente nesse relatório:

- *Análise de Volume de Transações por Agência*

Identificou quais agências estão gerando o maior volume de transações. A visualização do resultado foi através de gráfico de barras, que mostra o volume de transações por agência. Esta análise ajuda a identificar as agências mais movimentadas, permitindo alocar recursos de forma mais eficiente e identificar oportunidades de expansão.

- *Receita Mensal*

Monitora a receita mensal ao longo do tempo contemplado nas planilhas. Foi plotado um gráfico de linha mostrando a receita mensal que ajuda a identificar tendências sazonais, prever fluxos de caixa e ajustar estratégias financeiras e de marketing.

- *Volume de Transações por Dia da Semana*

Para responder as perguntas do item 3, plotei um gráfico de pizza mostrando o volume de transações por dia da semana em que se é possível identificar quais dias da semana têm maior volume de transações. Essa análise ajuda a otimizar o atendimento ao cliente e o planejamento de recursos, especialmente em dias de pico.

- *Valor Médio Movimentado por Dia da Semana*

Identifica quais dias da semana têm maior valor médio movimentado através de um gráfico de pizza para cada dia. Esse gráfico informa o planejamento de ofertas e promoções em dias específicos da semana para maximizar a receita.

- *Valores Movimentados no Início e no Final do Mês*

Determina se os maiores valores são movimentados no início ou no final do mês com intuito de ajustar o fluxo de caixa e a planejar campanhas promocionais em períodos de maior movimentação financeira. Esse resultado foi apresentado em um gráfico de pizza mostrando os valores movimentados no início e no final do mês.

Essas análises fornecem sugestões que podem ajudar na otimização das operações, no planejamento estratégico e na personalização das ofertas de produtos e serviços. Com essas informações, o BanVic pode tomar decisões mais assertivas, melhorar a experiência do cliente e aumentar a eficiência operacional.

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

Os resultados das análises realizadas mostram padrões e questões importantes que podem ajudar o BanVic a otimizar suas operações, melhorar a experiência dos clientes e aumentar a eficiência operacional.

- *Volume de Transações por Agência:* Algumas agências geram um volume significativamente maior de transações em comparação com outras. A partir dessa constatação, pode-se recomendar aos stakeholders a alocação mais recursos (pessoal, infraestrutura) para as agências mais movimentadas para melhorar o atendimento ao cliente e reduzir tempos de espera. Além disso, pode-se considerar a abertura de novas agências em regiões próximas a agências já movimentadas para capturar mais mercado.
- *Receita Mensal:* A receita mensal apresenta variações ao longo do tempo, com picos em determinados períodos. Utilizar as tendências de receita para prever fluxos de caixa e planejar melhor os investimentos pode ser uma das sugestões propostas a partir das análises. Uma sugestão para lidar com as variações sazonais é desenvolver campanhas de marketing para capitalizar nos períodos de alta receita e impulsionar as vendas em períodos de baixa.
- *Distribuição de Clientes por Tipo de Cliente:* A base de clientes é composta por diferentes tipos (PF, PJ). É possível personalizar produtos e serviços para atender às necessidades específicas de cada segmento de clientes, por exemplo, criando campanhas de marketing direcionadas para cada segmento, que aumentam a eficácia e a conversão.
- *Volume de Transações por Dia da Semana:* Os dias da semana apresentam variações no volume de transações. A partir dessa constatação, é possível ajustar a alocação de pessoal e recursos para os dias com maior volume de transações para garantir um atendimento eficiente e reduzir o tempo de espera dos clientes. Realizar promoções e ofertas especiais nos dias de menor movimento para equilibrar o fluxo de clientes ao longo da semana pode ser uma boa prática recomendada.
- *Valor Médio Movimentado por Dia da Semana:* Além do volume de transação flutuar ao longo da semana, certos dias têm um valor médio movimentado mais

alto do que outros. Planejar ofertas especiais e promoções nos dias com maior valor médio movimentado para pode contribuir para maximizar a receita. Essa prática pode ser desenvolvida ajustando as estratégias de marketing e vendas.

- *Valores Movimentados no Início e no Final do Mês:* O valor movimentado no início do mês é, em média, 10,26% maior do que no final do mês. É possível ajustar o fluxo de caixa para garantir a disponibilidade de recursos no início do mês, quando a movimentação financeira é maior, além de planejar campanhas promocionais nesse período para aproveitar a maior movimentação financeira e impulsionar as vendas.

Para Sofia Oliveira, CEO do BanVic, podemos recomendar – a partir da análise dos dados de ERP e CRM -, investir na melhoria das agências mais movimentadas para garantir um atendimento eficiente e reduzir os tempos de espera dos clientes, criar campanhas de marketing alinhadas com as tendências sazonais de receita para maximizar o impacto e a conversão, personalizar produtos e serviços de acordo com os diferentes segmentos de clientes para atender melhor às suas necessidades e aumentar a satisfação do cliente, ajustar a alocação de pessoal e recursos para os dias com maior volume de transações para garantir um atendimento eficiente, implementar um processo contínuo de análise de dados para monitorar o desempenho dos KPIs e ajustar as estratégias conforme necessário.

Essas recomendações podem ajudar o BanVic a tomar decisões mais assertivas, melhorar a experiência do cliente e aumentar a eficiência operacional, contribuindo para o crescimento e sucesso da empresa.

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

Busquei trabalhar com Python, linguagem para análise de dados em que tenho maior experiência. Todos os arquivos e scripts podem ser acessados através do repositório BANVIC_LIGHHOUSE_INCIDIO no meu GitHub:

https://github.com/0natalia/BANVIC_LIGHHOUSE_INCIDIO.git

Usei bibliotecas básicas e abertas para a construção do meu script, além do auxílio de inteligências artificiais open access para me ajudar em sua elaboração mais rápida e eficaz.