

Descrição de alterações e análises

Primeiramente, foi efetuada uma validação da normalização de cada um dos arquivos, a partir de uma revisão visual e estatística, verificando possíveis valores nulos ou repetidos e assegurar que as variáveis estão dentro das escalas previstas. Em seguida, converteu-se a coluna de datas para o formato `datetime`, garantindo a manipulação correta dos dados temporais.

Foi então criada uma dimensão de datas abrangente (*dim_dates*) para o período de 2010 a 2030, adicionando colunas adicionais como ano, mês, dia, dia da semana e outros detalhes temporais. Para associar cada transação a informações temporais detalhadas, foram unificados os dados de transações com a dimensão de datas, agrupando os dados por data, calculando o volume de transações por dia. Assim, calculou-se a média diária de transações, fornecendo uma visão clara do comportamento transacional diário.

Também foram realizadas análises para calcular a frequência e o tipo de transações mais comuns.

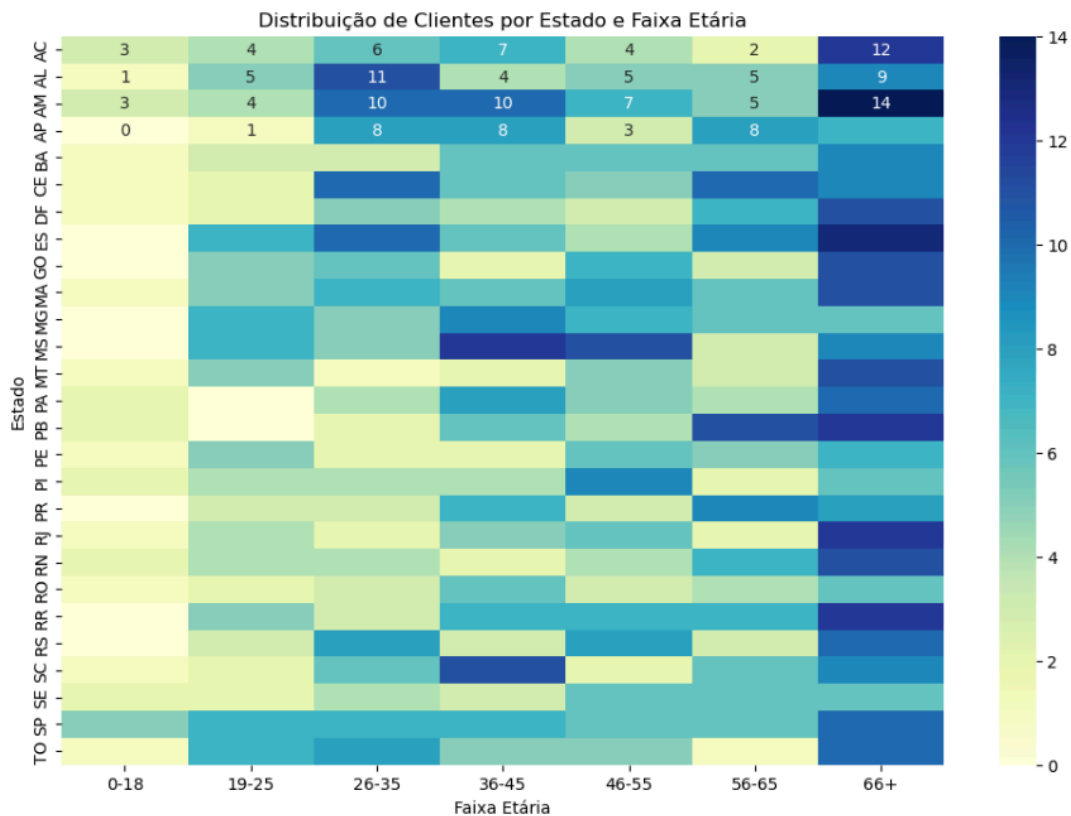
Para facilitar análises posteriores, a coluna de data de nascimento foi convertida para o formato `datetime`, permitindo cálculos diretos de idade com base na data atual. Com a idade calculada para cada cliente, o próximo passo foi agrupá-los em faixas etárias específicas (0-18; 19-25; 26-35; 36-45; 46-55; 56-65; 66+). A partir dos endereços dos clientes, foi então extraído o Estado de residência.

Com os clientes categorizados por faixa etária e estado, foi possível contabilizar quantos clientes pertencem a cada combinação única de estado e faixa etária. Por exemplo, quantos clientes estão na faixa etária de 26-35 anos no estado de São Paulo. Para tornar a análise mais clara e visualmente informativa, foi gerado um heatmap, que ajudou a visualizar facilmente a distribuição dos clientes entre os estados e faixas etárias. Cada célula no heatmap representa o número de clientes em uma determinada faixa etária e estado, com cores diferentes indicando diferentes níveis de contagem.

Esse processo foi realizado utilizando funções da própria biblioteca Pandas, tornando o processo eficiente e preciso.

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? Apresente sua análise exploratória realizada nos dados apresentados e cite quais perguntas de negócio podemos responder a fim de mostrar o valor que dados possuem para a empresa. Pergunta de nível teórico. O seu projeto final não necessariamente precisa trazer os mesmos ou todas as análises/perguntas de negócio citados nessa entrega.

A análise de dados demográficos e comportamentais permite criar perfis detalhados dos clientes, facilitando campanhas direcionadas e personalizadas.



Um exemplo seria o gráfico acima mostrando a distribuição de clientes nos estados do país e no Distrito Federal, com o objetivo de avaliar a distribuição de agências no país. Dessa forma, podemos observar em quais regiões do país o BanVic tem mais clientes, especialmente clientes idosos, que costumam preferir

atendimento em em agências físicas. Essas informações permitem que sejam analisadas a possibilidade de abertura de novas agências em diferentes áreas, ou mesmo o fechamento de agências obsoletas.

Entender a frequência e os tipos de transações mais comuns pode ajudar a otimizar os serviços oferecidos e identificar oportunidades para novos produtos financeiros.

nome_transacao	
Compra Crédito	24977
Compra Débito	17662
Pix - Realizado	13164
Pix - Recebido	5134
Depósito em espécie	4866
Saque	2879
TED - Recebido	1005
Pagamento de boleto	452
Transferência entre CC - Crédito	441
Estorno de Debito	403
DOC - Recebido	402
Transferência entre CC - Débito	303
TED - Realizado	130
DOC - Realizado	91
Pix Saque	90

No gráfico acima, percebe-se que o Pix é bastante popular, mesmo sendo uma tecnologia recente. O que sugere que os clientes do banco estão dando preferência a pagamentos instantâneos. Somando às compras no débito, é possível perceber que há uma tendência dos clientes darem preferência a realizar o pagamento por inteiro, evitando assim dívidas. No entanto, o número alto de transações de crédito pode sugerir uma oportunidade de oferecer, para clientes específicos, produtos financeiros adicionais, como cartões de crédito ou empréstimos, aumentando a liquidez do banco.

Dessa forma, é possível perceber que um modelo de segmentação de dados mais preciso possibilitará a criação de campanhas direcionadas ao público que buscamos atingir, aumentando a relevância e eficácia dos projetos e, por fim, resultando em maior retorno sobre investimento (ROI). Além de identificar padrões de comportamento e fatores de risco, permitindo a criação de estratégias que previnam perdas e danos, enquanto otimizam os serviços oferecidos.

Os dashboards e relatórios interativos se tornam então ferramentas que possibilitam uma visão mais clara e rápida de informações essenciais, facilitando a tomada de decisões estratégicas, baseando-as em dados. Os insights então

orientam o desenvolvimento de novos produtos e serviços que atendem às necessidades específicas de diferentes segmentos de clientes.

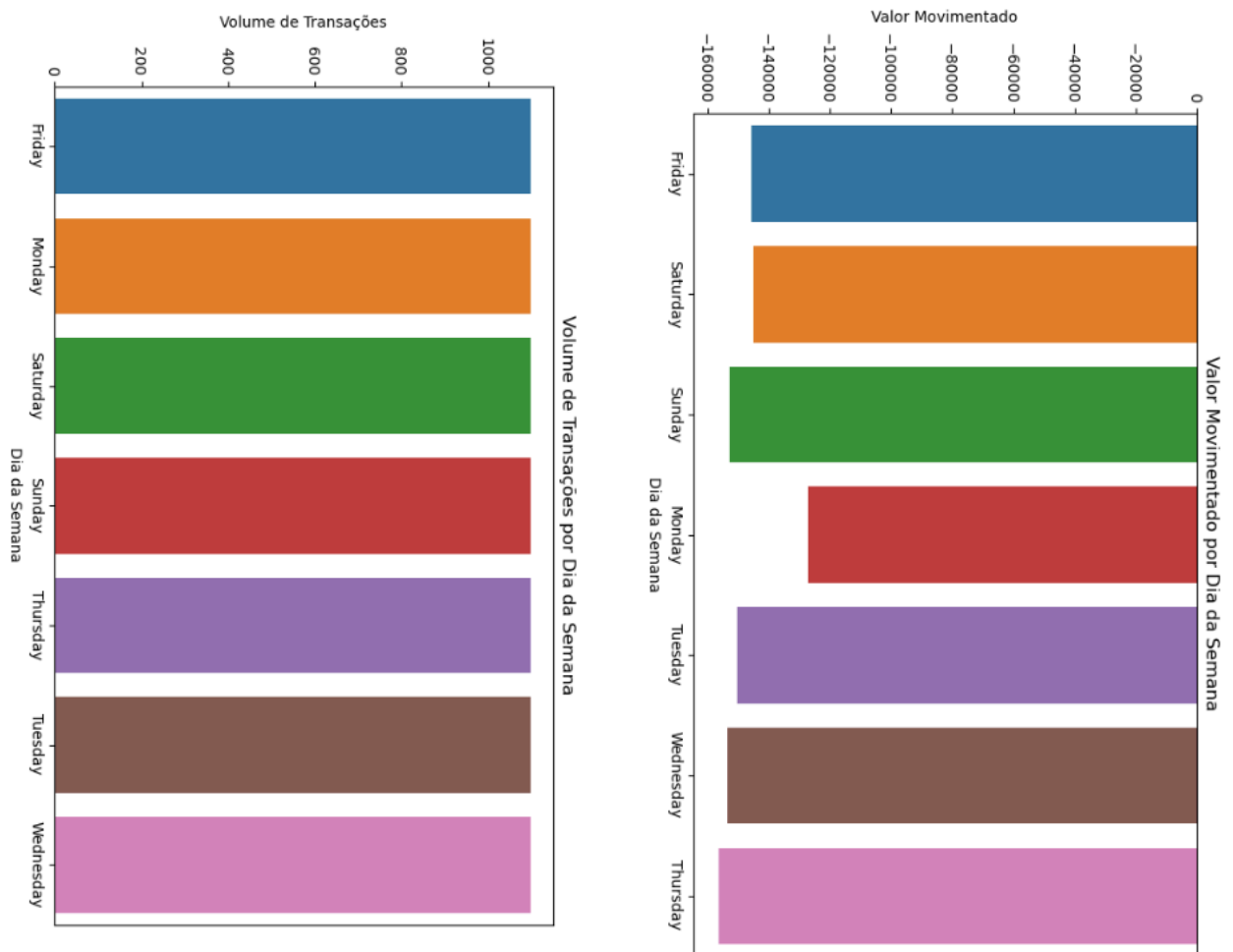
Implementar uma frente de Business Intelligence no BanVic, então proporcionará uma vantagem competitiva significativa, promovendo decisões mais informadas, aumentando a eficácia das campanhas de marketing, otimizando os serviços oferecidos e melhorando a experiência do cliente. De forma que a adoção dessa medida em longo prazo garantirá que as estratégias permaneçam relevantes e eficazes ao longo do tempo, possibilitando o crescimento sustentável da empresa.

Algumas das possíveis perguntas de negócio que poderiam ser respondidas utilizando os dados existentes incluem:

- I. Verificar a taxa de inclusão de novos clientes ao longo do tempo. Permitindo verificar quais campanhas obtiveram sucesso, ou procurar outros motivos que possam ter contribuído para o influxo de clientes.
 - II. Quais são os canais preferidos pelos clientes para realizar transações bancárias, e como podemos personalizar serviços para aumentar a satisfação do cliente?
 - III. Como podemos melhorar a eficiência operacional das agências, reduzindo o tempo médio de atendimento ao cliente?
 - IV. Como as variações nas taxas de juros afetam a demanda por produtos de crédito?
-
2. Quais serão os indicadores ou análises utilizadas para responder às perguntas de negócio citadas anteriormente? Pergunta de nível teórico. O seu projeto final não necessariamente precisa trazer os mesmos ou todos os indicadores/análises citados nessa entrega.
- I. Qual a taxa de inclusão de novos clientes ao longo do tempo?
 - A. Contagem de novos clientes ao longo dos anos, verificando a variação percentual no número de novos clientes em comparação com períodos anteriores.
 - II. Quais são os canais preferidos pelos clientes para realizar transações bancárias, e como podemos personalizar serviços para aumentar a satisfação do cliente?

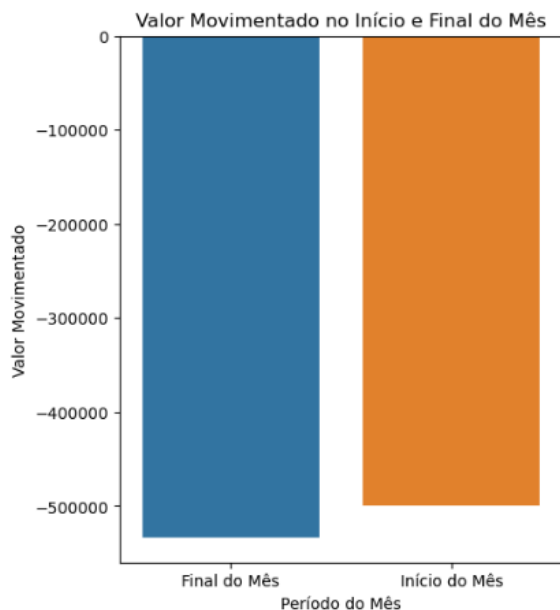
- A. Análise do número e valor das transações realizadas em diferentes canais (agências físicas, internet banking) e coleta de feedback dos clientes sobre a experiência em diferentes canais.
- III. Como as variações nas taxas de juros afetam a demanda por produtos de crédito?
 - A. Comparando a taxa de aprovação de propostas em diferentes períodos de variação das taxas de juros e analisando o número de solicitações de crédito recebidas durante períodos de alta e baixa das taxas de juros.
- 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. Pergunta prática e teórica.
 - Qual dia da semana tem, em média, maior volume de transações e qual tem, também em média, maior valor movimentado?

Como é possível observar nos gráficos abaixo, não há diferença considerável no volume de transações em diferentes dias da semana. Foi observado que o volume de transações realizadas permanece estável em todos os dias da semana. Em contrapartida, há uma leve distinção na média de valores movimentados, com a segunda-feira tendo o menor saldo de valores movimentados e a quinta-feira o maior.



- O BanVic tem, em média, os maiores valores movimentados no início ou final de mês? (Considere início do mês como sendo os primeiro 15 dias e o final do mês sendo os últimos 15 dias de cada mês).

Foi observado no gráfico abaixo, que o início do mês tem os menores valores movimentados durante o mês, e que o clientes do banco costumam movimentar maiores somas durante os últimos 15 dias do mês.



Além das análises propostas aqui, também seria interessante verificar o volume de transações por mês e trimestre para identificar atividades sazonais. Isso pode ajudar a entender se há sazonalidade nos comportamentos de transação dos clientes. Ou observar o volume e o valor das transações mês a mês para identificar tendências de crescimento ou declínio. Outra possibilidade seria detectar e investigar dias específicos com volumes de transações significativamente mais altos ou mais baixos que a média. Isso pode ajudar a entender os fatores que influenciam esses *outliers*, como eventos especiais ou promoções.

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.

Partindo do pressuposto que André Tech gostaria de indicadores econômicos, financeiros, demográficos e socioeconômicos, inicialmente seriam integradas dados do Instituto Brasileiro de Geografia e Estatística (IBGE), do Instituto de Pesquisa Econômica Aplicada (Ipea) e do Banco Central do Brasil (BC). Essas instituições registram altos níveis de confiabilidade, além de possuírem uma alta gama de variedade de dados disponíveis através do Portal Brasileiro de Dados Abertos

Inicialmente com foco na expansão da base de clientes, seria necessário aprofundar o conhecimento sobre a população brasileira, sendo essencial utilizar dados demográficos, econômicos e sociais facilitando o entendimento das necessidades dos diferentes segmentos da população. Por exemplo, campanhas direcionadas a jovens adultos em áreas urbanas podem ser ajustadas de maneira diferente das destinadas a idosos em áreas rurais, levando em consideração suas preferências e comportamentos financeiros.

Além disso, informações socioeconômicas, como níveis de renda média por região e taxas de escolaridade, oferecem insights valiosos sobre o poder de compra e o comportamento de consumo das diferentes populações. Isso é crucial para desenvolver produtos financeiros que sejam acessíveis e atraentes para o público-alvo do BanVic, promovendo uma maior adesão e satisfação entre os clientes.

Para integrar os dados ao banco de dados do BanVic seria necessário criar um ETL robusto e escalável. Começando com o Apache Spark para extração de dados diretamente da fonte, e então processar e transformar os dados, garantindo que eles sigam os padrões do banco. Para garantir maior eficiência, o Kubernetes é utilizado para gerenciar a orquestração dos containers do Spark. Por fim, utilizando Apache Airflow é possível agendar a execução das tarefas, monitorando o progresso do fluxo de trabalho e gerenciar automaticamente falhas ou reinicializações de tarefas, assegurando confiabilidade e consistência.

Para a CEO, Sofia Oliveira

Sua visão de que a análise de dados é crucial para elevar o Banco Vitória S.A. a novos patamares está bem alinhada com as demandas atuais do mercado. Compreender as operações e comportamentos dos clientes deve ser a base para melhorias significativas nos serviços de um banco.

Ficam aqui minhas recomendações para nossa futura parceria:

1. Implementação de uma Estratégia de Customer Insights: É crucial a criação de uma estratégia para capturar e analisar dados detalhados sobre o comportamento dos clientes em todos os canais de interação. Esta abordagem permitirá entender melhor as necessidades individuais dos clientes, além de possibilitar a antecipação de suas expectativas futuras. Utilizando técnicas avançadas de análise de dados, podemos identificar padrões e tendências de comportamento que não seriam perceptíveis de outra forma.
2. Personalização de Serviços e Ofertas: Ao utilizar insights detalhados, podemos adaptar serviços e ofertas de forma precisa às preferências individuais de cada cliente. Isso aumenta a relevância das comunicações, e melhora significativamente a experiência do cliente, promovendo uma maior fidelização e engajamento.
3. Fortalecimento da Cultura Data-Driven: Promover uma cultura organizacional que valorize e utilize dados como um ativo estratégico é essencial. Investir em educação e treinamento contínuos para capacitar os colaboradores em habilidades analíticas avançadas garantirá que todos na organização estejam alinhados com essa visão data-driven.

Essas recomendações e muitas outras serão os pilares que sustentarão a posição competitiva do Banco Vitória, e assegurarão o seu posicionamento como um banco inovador e confiável. Estou comprometida em contribuir com minhas habilidades e experiência em análise de dados para realizar essa visão. Juntas poderemos transformar insights em ações que impulsionam o sucesso contínuo do Banco Vitória.

Maíra Monteiro

Analista de Dados

Para mais informações:

https://github.com/MairaMonteiro/LH_EA-AD_MAIRAMONTEIRO

<https://www.linkedin.com/in/maira-monteiro>