**Documentação do Processo de Construção do Dashboard e Modelagem de Dados**

**1. Introdução**

Este documento detalha as decisões tomadas durante a modelagem dos dados, desenvolvimento do pipeline ETL e construção do dashboard para análise da performance financeira da empresa. O objetivo do projeto foi integrar múltiplas fontes de dados, estruturar um Data Warehouse eficiente e fornecer insights estratégicos por meio de um dashboard interativo e análises preditivas. A estratégia adotada permitiu consolidar dados dispersos e garantir que todas as informações críticas estivessem acessíveis para tomada de decisão.

Além disso, a abordagem priorizou a construção de um sistema escalável, garantindo que novos dados pudessem ser integrados sem comprometer o desempenho das consultas e visualizações. O processo de modelagem de dados também foi pensado para permitir análises aprofundadas de padrões de compra e comportamento do cliente. A estrutura final possibilitou insights valiosos sobre receita, segmentação de clientes e padrões sazonais, contribuindo para uma tomada de decisão mais assertiva.

**2. Modelagem de Dados e Pipeline ETL**

**2.1 Estruturação do Banco de Dados**

Optamos pelo **PostgreSQL** como banco de dados devido à sua robustez, escalabilidade e suporte nativo a operações analíticas. A modelagem seguiu um formato **dimensional**, utilizando o esquema **estrela**, pois facilita consultas analíticas de alto desempenho e melhora a experiência dos usuários de negócio ao explorar os dados. Além disso, a implementação de índices e particionamento ajudou a garantir eficiência na execução das consultas.

* **Fato:** Tabela transacoes armazenando movimentações financeiras dos clientes.
* **Dimensões:**
  + clientes: informações demográficas e de perfil de clientes.
  + produtos: categorias e valores dos produtos financeiros.
  + tempo: granularidade temporal para facilitar análises sazonais.

A estrutura do banco foi desenhada para garantir eficiência no acesso aos dados e flexibilidade na análise. Dessa forma, conseguimos facilitar a obtenção de insights estratégicos para diferentes setores da empresa, desde marketing até planejamento financeiro.

**2.2 Construção do Pipeline ETL**

Para a ingestão, transformação e carga dos dados, utilizamos **Python** com **pandas** e **SQLAlchemy** para manipulação dos dados e carga no PostgreSQL. As etapas do ETL foram cuidadosamente planejadas para garantir confiabilidade e escalabilidade.

1. **Extração:**
   * Dados brutos foram carregados a partir de múltiplas fontes e integrados ao pipeline.
   * Definição de processos automatizados para futuras atualizações.
   * Validação da consistência dos dados antes da transformação.
2. **Transformação:**
   * Tratamento de valores nulos e duplicados para garantir integridade.
   * Normalização e padronização dos dados para consistência.
   * Cálculo de métricas agregadas, como total gasto e frequência de compras.
   * Aplicação de regras de negócio para derivação de novos indicadores.
3. **Carga:**
   * Inserção dos dados transformados no PostgreSQL no schema desafio\_bi.
   * Implementação de particionamento e índices para otimização de consultas.
   * Validação de integridade dos dados pós-carga.

Essas decisões garantiram que o pipeline fosse eficiente, escalável e capaz de suportar análises avançadas, tornando a estrutura resiliente e adaptável a mudanças futuras.

**3. Construção do Dashboard**

**3.1 Ferramenta Escolhida**

A escolha do **Metabase** foi baseada na sua capacidade de integração com o PostgreSQL e sua interface intuitiva para usuários não técnicos. Além disso, o Metabase oferece filtros interativos, dashboards dinâmicos e baixa complexidade na implementação. Outra vantagem do Metabase é sua flexibilidade para permitir análises ad hoc, onde os usuários podem explorar os dados sem necessidade de intervenção técnica.

**3.2 Critérios para Escolha das Métricas e Visualizações**

A definição das métricas do dashboard foi guiada pelos seguintes princípios:

* **Monitoramento da Performance**:
  + Receita total e volume de transações.
  + Receita por produto e por região.
  + Evolução da receita ao longo do tempo.
* **Análises Avançadas:**
  + Identificação de padrões sazonais e tendências de vendas.
  + Segmentação de clientes com base no comportamento de compra.
  + Análise de churn para prever perda de clientes.

Cada visualização foi estruturada para permitir que tomadores de decisão identifiquem rapidamente oportunidades e riscos no desempenho financeiro da empresa, garantindo uma abordagem baseada em dados.

**4. Análise Preditiva e Insights Estratégicos**

**4.1 Modelagem Preditiva**

Para prever tendências de vendas e identificar padrões de churn, utilizamos **Facebook Prophet** para modelagem de séries temporais e **Random Forest** para classificação de clientes em churn e não churn. Ambos os modelos foram escolhidos devido à sua eficácia e capacidade de gerar previsões precisas.

**Principais Insights e Resultados Encontrados**

1. **Tendência Geral de Receita:**
   * O modelo de séries temporais revelou uma **leve tendência de queda** na receita ao longo do tempo.
   * Identificamos picos sazonais de vendas em **Maio, Julho e Setembro**, e uma queda significativa em **Novembro e Dezembro**.
2. **Segmentação de Clientes:**
   * Utilizamos **K-Means Clustering** para agrupar clientes em três perfis:
     1. **Clientes VIP**: Alta frequência de compra e alto ticket médio.
     2. **Clientes Ocasionalmente Ativos**: Compram de forma esporádica.
     3. **Clientes de Baixo Engajamento**: Risco alto de churn.
   * Os clusters foram utilizados para personalizar campanhas de marketing e aumentar retenção.
3. **Fatores de Churn:**
   * A variável mais impactante na previsão foi **Recência (tempo desde a última compra)**, seguida pelo **Total Gasto**.
   * Identificamos que **clientes que não compram há mais de 420 dias têm alta probabilidade de churn**.
   * Foi sugerida uma estratégia de reengajamento com promoções e comunicação segmentada.

**5. Conclusão e Próximos Passos**

**Decisões Estratégicas Baseadas nos Insights**

✔ Criar **campanhas de retenção** focadas em clientes com alta recência e baixo engajamento. ✔ Implementar **programas de fidelidade** para clientes de alto valor. ✔ Ajustar ofertas promocionais para aumentar a recorrência de compras. ✔ Melhorar a segmentação de campanhas com base na análise de clusters. ✔ Acompanhar a evolução das métricas de churn e engajamento. ✔ Aplicar estratégias sazonais para maximizar vendas em períodos de alta demanda.

**Possíveis Melhorias Futuras**

* Incorporar **dados macroeconômicos** para enriquecer a previsão de vendas.
* Desenvolver um modelo de **Lifetime Value (LTV)** para priorizar investimentos nos clientes mais lucrativos.
* Automatizar o pipeline ETL utilizando ferramentas como **Airbyte, dbt e Dagster**.
* Criar um sistema de notificações para alertar sobre clientes com alto risco de churn.

Este documento serve como guia para futuras melhorias e como referência para a tomada de decisão baseada em dados, garantindo um fluxo contínuo de insights estratégicos e permitindo a evolução constante das análises empresariais.