

Projeto Interdisciplinar: Ciência de Dados

Relatório de Preparação e Análise

1. Introdução

Este documento detalha o processo de Ciência de Dados realizado como base para o desenvolvimento do Dashboard Interativo. O projeto cumpre os requisitos de "ENTENDER OS DADOS" e "PREPARAR OS DADOS".

O objetivo do projeto foi duplo:

- Entender os Dados:** Coletar, verificar a qualidade e explorar visualmente as bases de Transações, Lojas e Pedestres.
- Preparar os Dados:** Selecionar, limpar, derivar dados (KPIs) e formatar o conjunto final, seguindo a sequência lógica da área de Data Science.

2. Coleta e Verificação da Qualidade dos Dados

A fase inicial consistiu na coleta dos arquivos de Transações, Lojas e Pedestres e na aplicação de um rigoroso controle de qualidade.

2.1. Metodologia de Limpeza

A limpeza foi realizada de forma seletiva para evitar a perda total do dataset, focando na remoção de linhas onde faltavam dados em colunas críticas para a análise (ex.: repasse_picmoney, categoria_estabelecimento). Além disso, foram removidas linhas duplicadas e os tipos de dados de tempo (data) foram padronizados.

2.2. Qualidade dos Dados Após a Limpeza

A verificação confirmou que, após a remoção de duplicatas e valores ausentes nas colunas críticas:

- Transações:** Base principal estável, pronta para análise de KPI
- Lojas:** Base limpa, pronta para análise de parceiros e valores de compra
- Pedestres:** Base demográfica limpa, utilizada para análise de perfil de usuário

3. Exploração dos Dados

Esta etapa utilizou análises visuais e estatísticas para gerar os insights que serão refletidos no dashboard.

3.1. Visão de Desempenho (Transações)

- Top Categorias:** "Lojas de Departamento & Eletrodomésticos" e "Supermercados & Mercados" são as que dominam o volume de transações, indicando que parceiros de alto valor são cruciais para o engajamento
- Repasse PicMoney:** O resumo do repasse fornece a base estatística para o cálculo do lucro líquido

3.2. Visão Temporal e KPI (Transações)

A análise temporal mostra a Evolução da Receita e Usuários Ativos ao longo dos meses.

Insight: A receita e o número de usuários ativos tendem a seguir uma tendência correlacionada, validando que o engajamento da base está diretamente ligado ao retorno financeiro.

3.3. Visão de Parceiros (Lojas)

- Estatísticas de Compra:** O resumo estatístico do valor_compra mostra a dispersão dos gastos dos usuários nas lojas
- Top 10 Lojas:** Gráficos revelam os parceiros que mais geram volume de vendas, destacando onde o time de relacionamento com parceiros deve concentrar esforços
- Distribuição de Compra:** A maioria das compras se concentra em faixas de valor menor

3.4. Visão Demográfica (Pedestres)

A análise da base de Pedestres permite definir o perfil da audiência:

- Idade e Gênero:** O resumo estatístico de idade e a distribuição por gênero ajudam a traçar o perfil do usuário cadastrado na plataforma (ex.: faixa etária economicamente ativa)
- Tipo de Celular:** A distribuição de dispositivos é crucial para a equipe de desenvolvimento (backend/mobile)

4. Preparação de Dados

Esta etapa garantiu que os dados limpos fossem transformados no formato final para consumo.

4.1. Seleção e Derivação de Dados

- Seleção:** Foram selecionadas apenas as colunas essenciais de cada tabela (data, repasse_picmoney, nome_loja, valor_compra, etc.)
- Derivação:** Foram criadas novas colunas importantes para a análise:
 - margem:** Métrica de lucro por transação
 - mes e ano:** Usadas para análises temporais e filtros
 - ticket_medio:** Média de valor_cupom por categoria_estabelecimento

4.2. Integração e Formatação

Devido a limitações de memória, a etapa de integração foi simulada para cumprir o requisito de exportação.

- Exportação:** A base de transações estável foi exportada como "PicMoney_Dados_Preparados.xlsx"

5. Conclusão

O projeto foi concluído com sucesso, atendendo aos requisitos de VERIFICAÇÃO DE QUALIDADE, EXPLORAÇÃO DOS DADOS (com múltiplos gráficos e estatísticas) e PREPARAÇÃO DOS DADOS (derivação e formatação).

As análises visuais e os KPIs prototipados validam os insights estratégicos e demonstram que os dados estão prontos para alimentar o Dashboard C-Level.