

1. Introdução

Este documento detalha o processo de Ciência de Dados realizado como base para o desenvolvimento do Dashboard Interativo da PicMoney. O objetivo é duplo:

1. **Preparar os Dados:** Documentar as etapas de derivação e integração necessárias para transformar os dados limpos (`_cleaned.csv`) em um conjunto de dados final e pronto para consumo pelo backend.
2. **Analisar os Dados:** Conduzir uma análise exploratória visual (prototipada em Colab) que revela os principais KPIs e *insights* que serão apresentados no dashboard C-Level.

Este trabalho cumpre os requisitos de "PREPARAR OS DADOS" (Derivar, Integrar, Formatar) e "ENTENDER OS DADOS" (Explorar) definidos no escopo do projeto.

2. Preparação e Integração de Dados

Partindo dos arquivos `_cleaned.csv`, foram necessárias etapas adicionais de derivação e integração para criar o dataset final que alimenta o dashboard (`transacoes_final_integrada.csv`).

2.1. Derivação de Dados

Embora os dados estivessem limpos, ao serem carregados a partir dos CSVs, os tipos de dados precisavam ser "derivados" (convertidos) para formatos corretos para análise:

- As colunas `data` (nos arquivos `transacoes_cleaned.csv` e `pedestres_cleaned.csv`) foram convertidas de *string* para o formato `datetime`, permitindo análises temporais.
- A coluna `idade` (em `players_cleaned.csv`) foi convertida para o tipo *numérico* para possibilitar cálculos estatísticos.

2.2. Integração dos Dados

A necessidade mais crítica do dashboard era a capacidade de filtrar dados financeiros por características demográficas dos usuários. Para isso, os dados foram integrados:

- **Chave de Junção:** Foi utilizada a coluna `celular` como chave primária.
- **Processo:** Foi realizado um `LEFT JOIN` (merge) entre a base principal `transacoes_cleaned.csv` (100.000 transações) e a base `players_cleaned.csv` (10.000 players).
- **Resultado:** Foi gerado o arquivo `transacoes_final_integrada.csv`, onde cada transação agora possui as colunas demográficas (`idade`, `sexo`, `Dispositivo`) do player que a realizou.

Com este dataset final, o backend pode realizar agregações complexas de forma eficiente, sem a necessidade de joins em tempo real.

3. Análise Exploratória e Visual (Prototipagem do Dashboard)

Com os dados integrados, foram geradas as seguintes análises visuais (prototipadas em Colab usando Plotly) para validar as visões do dashboard:

3.1. Visão Geral: KPIs Estratégicos

A análise agregada dos dados revela os principais indicadores de performance da plataforma:

- **Receita Bruta Total (Valor de Cupom):** R\$ 49.992.673,36
- **Comissão PicMoney (Receita Líquida):** R\$ 2.417.804,41
- **Total de Transações:** 100.000
- **Clientes Únicos:** 9.986 (indicando baixa recorrência no período)
- **Ticket Médio (por Cupom):** R\$ 499,93

3.2. Visão Temporal: Receita ao Longo do Tempo

O gráfico de linha da receita ao longo do período de análise (Julho de 2025) mostra uma distribuição relativamente estável, com picos de receita e comissão em dias específicos, sugerindo a necessidade de investigar quais campanhas ou eventos ocorreram nessas datas.

3.3. Visão de Desempenho: Cupons e Categorias

A análise de desempenho de produtos revela:

- **Tipo de Cupom (Gráfico Donut):** Os cupons do tipo "Produto" são os mais capturados, seguidos por "Cashback" e "Desconto". Isso indica que a oferta de produtos específicos é um grande atrativo para os usuários.
- **Top Categorias (Gráfico de Barras):** As categorias "Lojas de Departamento & Eletrodomésticos" e "Supermercados & Mercados" são as que mais geram receita, dominando a análise. Isso sugere que os parceiros *high-ticket* são cruciais para a receita bruta.

3.4. Visão Geográfica: Densidade de Pedestres

O mapa de calor (heatmap) focado na Av. Paulista (usando pedestres_cleaned.csv) mostra uma alta concentração de pedestres perto de estações de metrô (Consolação, Brigadeiro) e grandes centros comerciais (Shopping Cidade São Paulo, Pátio Paulista).

- **Insight (CTO/CEO):** Essas são áreas de alta prioridade para garantir a cobertura de georreferenciamento e para focar a aquisição de novos parceiros comerciais.

3.5. Visão Demográfica: Análise de Usuários

A integração dos dados de players permitiu a criação de um gráfico de barras agrupado por faixa etária e gênero (o protótipo para a "pirâmide demográfica" do dashboard).

- **Principal Segmento:** A maior concentração de players (usuários cadastrados) está na faixa de **26 a 35 anos**, seguida pela faixa de **36 a 45 anos**.

- **Gênero:** Há uma participação maior de usuários do sexo "Masculino" e "Feminino" (com "Outro" sendo uma parcela menor), com uma distribuição relativamente equilibrada entre os dois principais gêneros nas faixas de maior volume.
- **Insight (CFO/CEO):** O perfil de usuário principal é economicamente ativo (26-45 anos). As campanhas de marketing e a aquisição de parceiros devem focar nos interesses desse grupo.

4. Conclusão

O processo de preparação e integração de dados foi concluído com sucesso, resultando em um dataset final (`transacoes_final_integrada.csv`) que permite a implementação de todas as visualizações e filtros complexos exigidos pelo dashboard C-Level.

As análises visuais prototipadas no Colab validam a lógica que será usada nas APIs do backend e já apontam *insights* estratégicos claros sobre o desempenho financeiro, geográfico e demográfico da PicMoney.