

# **Análise, Limpeza e Exploração de Dados da Empresa PicMoney**

## **1. Introdução**

O presente trabalho tem como objetivo descrever o processo de limpeza, tratamento e exploração dos dados da empresa PicMoney, uma startup do setor financeiro que se encontra em fase de pré-lançamento de seu aplicativo. A organização tem como propósito oferecer soluções inovadoras para o controle financeiro pessoal, integrando ferramentas de gestão de gastos, investimentos e metas econômicas em uma plataforma intuitiva e acessível.

Dada a natureza do empreendimento e a ausência de dados reais provenientes de usuários, a equipe de análise utilizou bases de dados simuladas para representar o comportamento esperado de clientes, transações financeiras e indicadores de desempenho interno. O estudo buscou aplicar técnicas de ciência de dados para validar a consistência das informações, identificar padrões relevantes e fornecer subsídios para decisões estratégicas da empresa.

A importância deste trabalho está relacionada à preparação de uma base de dados sólida e confiável antes do lançamento oficial do produto. A qualidade dos dados é um fator determinante para a eficácia dos algoritmos analíticos, das previsões de receita e da experiência dos usuários. Assim, o presente relatório apresenta de forma detalhada as etapas de tratamento e análise dos dados, bem como as principais conclusões obtidas.

## **2. Metodologia**

O processo de análise de dados foi dividido em três etapas principais: limpeza, tratamento e exploração dos dados. Todas as fases foram executadas em ambiente Python, utilizando bibliotecas amplamente aplicadas na área de ciência de dados, como *pandas*, *numpy*, *matplotlib* e *seaborn*.

### **2.1. Limpeza dos Dados**

A etapa de limpeza consistiu na identificação e correção de inconsistências nas bases simuladas da PicMoney. Inicialmente, foi realizada a verificação de valores ausentes (*missing values*) e duplicados, com posterior remoção ou imputação conforme o contexto de cada variável. Foram também aplicadas rotinas para uniformizar os tipos de dados (por exemplo, conversão de colunas de data e moeda), além da padronização de nomes de colunas e categorias.

Além disso, foram verificadas possíveis inconsistências lógicas, como transações com valores negativos indevidos, cadastros sem identificação de cliente e registros duplicados de movimentações. Esses ajustes foram fundamentais para garantir a integridade e a coerência das análises subsequentes.

### **2.2. Tratamento dos Dados**

Após a limpeza, foi conduzido o processo de tratamento e transformação dos dados, com o objetivo de prepará-los para análises exploratórias e modelagens futuras. Entre as principais ações realizadas, destacam-se:

- Criação de novas variáveis derivadas, como média de gasto mensal por cliente, índice de retenção e taxa de conversão de usuários simulados;
- Normalização de dados financeiros para permitir comparações entre diferentes períodos;
- Agrupamento e agregação de dados por perfis de clientes e categorias de despesas;
- Categorização de variáveis qualitativas em formatos numéricos para posterior análise estatística.

Essas transformações permitiram compreender melhor o comportamento dos usuários e estimar métricas de desempenho relevantes para as áreas financeira e de marketing da empresa.

### 2.3. Exploração e Visualização dos Dados

Com as bases tratadas, foi realizada uma análise exploratória de dados (EDA – Exploratory Data Analysis). Foram gerados gráficos e tabelas para investigar tendências, dispersões e correlações entre variáveis. A visualização dos dados desempenhou um papel essencial na interpretação dos resultados, permitindo identificar padrões como:

- Distribuição de renda e gastos dos clientes simulados;
- Categorias de despesa mais recorrentes;
- Comparativo de receita simulada por período;
- Projeções de crescimento baseadas em diferentes cenários de adesão ao aplicativo.

A etapa exploratória serviu como base para insights estratégicos sobre o comportamento do público-alvo e o potencial de mercado da PicMoney.

## 3. Resultados e Discussão

A análise dos dados simulados da PicMoney revelou informações relevantes para o planejamento de sua operação inicial. Observou-se que a maioria dos usuários simulados apresentava perfil de renda média, com concentração de gastos em categorias como alimentação, transporte e lazer.

A taxa de retenção projetada indicou um comportamento estável de fidelização, enquanto os indicadores financeiros demonstraram potencial de crescimento sustentável a médio prazo. A simulação também apontou possíveis gargalos na gestão de custos e oportunidades de otimização nas campanhas de aquisição de clientes.

Além disso, a exploração dos dados permitiu detectar tendências sazonais nos gastos e receitas, sugerindo que o aplicativo poderia se beneficiar de recursos analíticos baseados em aprendizado de máquina, como previsões de fluxo de caixa e recomendações personalizadas de economia.

De forma geral, o trabalho de limpeza e tratamento resultou em um conjunto de dados estruturado, coerente e pronto para futuras aplicações analíticas, contribuindo diretamente para o desenvolvimento do produto e a definição de suas estratégias iniciais.

#### **4. Conclusão**

O presente estudo apresentou o processo de limpeza, tratamento e exploração dos dados da empresa PicMoney, demonstrando a importância da preparação e consistência das informações antes do lançamento de uma solução digital baseada em dados.

A metodologia aplicada garantiu a integridade das bases simuladas e possibilitou análises exploratórias de alto valor para as áreas financeira e estratégica da organização. Os resultados obtidos evidenciaram padrões de comportamento relevantes e forneceram subsídios para a tomada de decisão em relação ao público-alvo, monetização e estratégias de retenção.

Em etapas futuras, recomenda-se a implementação de modelos preditivos e análises de machine learning com dados reais, a fim de aprimorar as estimativas e validar as hipóteses observadas durante a fase exploratória. A experiência obtida neste trabalho reforça a relevância da ciência de dados como ferramenta essencial para o sucesso de startups no contexto financeiro digital.