

FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO - FECAP

Bacharelado em Ciência da Computação

Projeto Interdisciplinar: Ciência de Dados – Professor Eduardo Savino Gomes

Estudantes:


- André Gregório dos Santos – RA: 24026489
- Guilherme Reis Fogolin de Godoy – RA: 24026241
- Pedro Henrique Nascimento Lemos – RA: 23025380
- Yan Ramos Cezareto – RA: 24026005

Turma: 4NACOMP_S

Entrega 01 - Projeto Interdisciplinar: Ciência de Dados

Análise dos dados fornecidos pela PicMoney

Objetivo: O objetivo deste relatório é analisar detalhadamente a base de dados da PicMoney, investigando o perfil e comportamento dos usuários (players), bem como o desempenho das transações e estabelecimentos. O estudo envolve desde a importação e verificação da qualidade dos dados até a exploração e extração de insights relevantes para apoiar decisões estratégicas.

 **Importante!** Devidos scripts e visões estão disponíveis no arquivo do Colab. Para acesso: [Entrega 01 - Professor Savino](#).

Coleta e Descrição dos Dados

Carregamento dos dados: A primeira etapa do projeto envolve importar os conjuntos de dados para o ambiente de análise (Google Colab). Para isso, utilizamos a biblioteca Pandas para ler os quatro arquivos CSV fornecidos (Players, Transações, Pedestres e Lojas), que servem como base para toda a análise.

Bases de dados:

Players: Contém informações cadastrais dos usuários do aplicativo, como idade, sexo e localização. Esses usuários são os “caçadores” de cupons.

Transações: Registra todas as transações de cupons capturados, incluindo dados sobre valores, estabelecimentos e datas.

Pedestres: Base simulada com informações de potenciais usuários em uma área específica (Avenida Paulista). Esses dados ajudam a entender o potencial de atração de novos players em regiões de grande movimento.

Lojas: Conjunto de dados de teste com informações sobre os estabelecimentos parceiros e os valores dos cupons disponibilizados.

Verificação da qualidade e limpeza dos dados

Objetivo da limpeza de dados: O objetivo é identificar e corrigir inconsistências, erros de formatação e valores ausentes, garantindo que a análise seja confiável e precisa.

O processo de limpeza inclui as seguintes etapas:

Padronização de categorias: Uniformizar nomes e categorias (como tipos de cupons ou estabelecimentos) para evitar divergências na análise.

Correção de tipos de dados: Ajustar colunas para os tipos corretos (por exemplo, datas, números e textos) para cálculos e operações precisas.

Limpeza de campos: Remover espaços em branco, caracteres especiais ou registros irrelevantes que possam atrapalhar a análise.

Tratamento de valores ausentes: Preencher ou excluir dados faltantes de forma adequada para não comprometer resultados estatísticos.

Correção de coordenadas: Ajustar ou validar informações geográficas (latitude e longitude) para análises espaciais corretas.

Exploração de dados

Resumo da análise dos dados da PicMoney

01. Entendimento dos dados: O conjunto de dados analisado é composto por quatro bases principais: Players, Transações, Pedestres e Lojas, cada uma contendo informações estratégicas sobre usuários, movimentações de cupons, potenciais clientes e estabelecimentos parceiros. O objetivo da análise é compreender o comportamento dos usuários, identificar padrões de consumo, e avaliar o desempenho do negócio.

1.1. Players: Base cadastral com informações pessoais dos usuários, como idade, sexo, cidade e bairro de residência, cidade e bairro de trabalho/escola, e categorias de estabelecimentos frequentados.

1.2. Transações: Registro de todas as transações realizadas pelos usuários, incluindo valor do cupom, tipo de cupom (Cashback, Produto ou Desconto), estabelecimento e categoria do estabelecimento.

1.3. Pedestres: Base simulada com dados de potenciais clientes em locais de alta circulação (ex: Avenida Paulista), contendo informações de localização, dispositivo móvel e histórico de cupons capturados.

1.4. Lojas: Base de teste com informações sobre estabelecimentos parceiros e valores dos cupons oferecidos.

02. Coleta e Preparação dos Dados: Os dados foram carregados no Google Colab a partir de arquivos CSV e submetidos a um processo de limpeza e padronização, que incluiu:

2.1. Conversão de datas para o formato adequado.

2.2. Remoção de caracteres especiais em números de celular.

2.3. Padronização de categorias de estabelecimentos e tipos de cupons.

2.4. Preenchimento de valores ausentes com “Não informado” ou valores neutros.

2.5. Correção de coordenadas geográficas com múltiplos pontos decimais.

Essa preparação garante a confiabilidade e consistência dos dados para análise.

03. Descrição dos dados: Após a limpeza, as principais características dos datasets são:

3.1. Players: 10.000 registros, média de idade 52,7 anos, com quase totalidade de dados cadastrais preenchidos, exceto cidades e bairros de trabalho/escola parcialmente ausentes.

3.2. Transações: 100.000 registros, tipos de cupons distribuídos de forma equilibrada entre Cashback, Produto e Desconto, com valores de cupom e repasse PicMoney totalmente informados.

3.3. Pedestres: 100.000 registros, dados de localização completos, porém parcialmente ausentes no histórico de compras.

3.4. Lojas: 10.000 registros, com informações completas sobre localização, tipo de loja e valores de compra e cupom.

4. Exploração dos Dados (EDA) e KPIs: Foram analisados os seguintes indicadores chave (KPIs):

4.1. Número de jogadores ativos: 9.999.

4.2. Número total de transações: 100.000.

4.3. Número de pedestres registrados: 99.935.

4.4. Número de lojas: 15

4.5. Distribuição de tipos de cupons:

- Cashback: 33.556.

- Produto: 33.328.

- Desconto: 33.116.

A análise exploratória visual revelou padrões de comportamento, como predominância de faixas etárias e categorias de estabelecimentos mais frequentados, e permitiu identificar áreas de oportunidade para captação de novos usuários.

05. Conclusão: O processo de limpeza e análise dos dados possibilitou um entendimento profundo do perfil dos usuários e do desempenho das transações. Os KPIs e visualizações fornecem subsídios para decisões estratégicas, como campanhas de marketing, expansão de parcerias e melhorias na experiência do usuário.

Além disso, a análise indica que a PicMoney possui uma base de usuários diversificada, com predominância em determinadas faixas etárias, o que permite segmentar campanhas de forma mais assertiva. O volume significativo de transações demonstra forte engajamento dos usuários com os cupons, enquanto a concentração das lojas em categorias estratégicas mostra que a empresa consegue gerar valor tanto para parceiros quanto para consumidores. Esses insights reforçam o potencial de crescimento da PicMoney, indicando oportunidades de expansão geográfica, diversificação de categorias de parceiros e otimização de campanhas de cashback e descontos, alinhando-se aos objetivos de aumentar a captação e retenção de usuários na plataforma.