

1. Descrição dos Dados

Para a realização deste projeto, foram disponibilizados quatro conjuntos de dados em formato CSV, que compõem o ecossistema de informações da PicMoney. A coleta inicial foi realizada através do carregamento desses quatro arquivos para um ambiente de análise em Python. Cada arquivo possui uma finalidade específica:

- **PicMoney-Base_Cadastral_de_Players:** Contém os dados demográficos dos usuários cadastrados no aplicativo. Serve como a principal fonte para entender o perfil do público, incluindo informações como idade, sexo e bairro de residência. O resumo da estrutura desta base é apresentado abaixo.

	tipo de dado	Valores Distintos	Valores Nulos	preenchido
celular	object	9999	0	100.00
data_nascimento	object	8346	0	100.00
idade	int64	76	0	100.00
sexo	object	3	0	100.00
cidade_residencial	object	530	0	100.00
bairro_residencial	object	21	0	100.00
cidade_trabalho	object	1	3969	60.31
bairro_trabalho	object	21	3969	60.31
cidade_escola	object	1	6932	30.68
bairro_escola	object	21	6932	30.68
categoria_frequentada	object	29	0	100.00

- **PicMoney-Base_de_Transacoes_-_Cupons_Capturados:** É o registro histórico completo de todas as interações de captura de cupons (100.000 linhas). Detalha qual usuário capturou qual cupom e em que data e horário. É a base primária para análises de comportamento e volume de transações.

	Tipo de Dado	Valores Distintos	Valores Nulos (NA)	% Preenchido
celular	object	4813	0	100.00
data	object	31	0	100.00
hora	object	65	0	100.00
nome_estabelecimento	object	33	0	100.00
bairro_estabelecimento	object	21	0	100.00
categoria_estabelecimento	object	29	0	100.00
id_campanha	object	9000	0	100.00
id_cupom	object	94680	0	100.00
tipo_cupom	object	3	0	100.00
produto	object	182	66672	33.33
valor_cupom	float64	60201	0	100.00
repasso_picmoney	float64	24041	0	100.00

- **PicMoney-Base_Simulada_-_Pedestres_Av__Paulista:** Uma base de dados simulada contendo a movimentação de pedestres em uma área geográfica estratégica. Contém dados de geolocalização que, em análises futuras, podem ser utilizados para otimizar a distribuição de cupons.

	Tipo de Dado	Valores Distintos	Valores Nulos (NA)	% Preenchido
celular	object	99935	0	100.00
data	object	1	0	100.00
horario	object	65	0	100.00
local	object	10	0	100.00
latitude	object	100000	0	100.00
longitude	object	100000	0	100.00
tipo_celular	object	2	0	100.00
modelo_celular	object	10	0	100.00
possui_app_picmoney	object	2	0	100.00
data_ultima_compra	object	30	40043	59.96
ultimo_tipo_cupom	object	3	40043	59.96
ultimo_valor_capturado	float64	34527	40043	59.96
ultimo_tipo_loja	object	8	40043	59.96
idade	int64	55	0	100.00
sexo	object	3	0	100.00

- **PicMoney-Massa_de_Testes_com_Lojas_e_Valores:** Um subconjunto de dados de transações mais detalhado (10.000 linhas). Sua principal característica é a inclusão da coluna `valor_compra`, tornando-a a fonte de dados essencial para todas as análises financeiras, como cálculo de receita e ticket médio.

	Tipo de Dado	Valores Distintos	Valores Nulos (NA)	% Preenchido
numero_celular	object	9999	0	100.0
data_captura	object	31	0	100.0
tipo_cupom	object	3	0	100.0
tipo_loja	object	8	0	100.0
local_captura	object	7	0	100.0
latitude	object	10000	0	100.0
longitude	object	10000	0	100.0
nome_loja	object	15	0	100.0
endereco_loja	object	14	0	100.0
valor_compra	float64	9498	0	100.0
valor_cupom	float64	6503	0	100.0

2. Verificação e Tratamento da Qualidade dos Dados

Após a coleta e a descrição inicial, foi executado um processo de verificação e limpeza para garantir a precisão e a confiabilidade da análise. As tabelas de resumo acima, por exemplo, revelaram a presença de valores nulos (NA) e tipos de dados (dtype) que precisavam de tratamento. As seguintes etapas foram realizadas via script em Python:

- **Conversão de Tipos de Dados:** Campos de data, como `data_nascimento` e `data_captura`, que estavam em formato de texto (object), foram convertidos para o formato datetime padrão. Esta etapa é fundamental para permitir cálculos e análises temporais corretas.
- **Tratamento de Valores Ausentes:** Colunas com dados faltantes, como `bairro_trabalho` na base de players (que possuía 3969 valores nulos), foram preenchidas com um valor padrão ("Não Informado") para preservar a integridade dos registros para outras análises.
- **Limpeza e Padronização:** Foram aplicadas correções pontuais, como a padronização de nomes de colunas (`numero_celular` para `celular`), para permitir a unificação correta das diferentes bases de dados.

3. Exploração dos Dados e Análise de Relações

Com a base de dados tratada, a análise exploratória foi conduzida para "ENTENDER OS DADOS", conforme o objetivo desta entrega. Foram geradas agregações numéricas para extrair insights sobre o comportamento dos clientes e a performance financeira.

- **Relação: Gasto Médio por Compra (Ticket Médio) por Faixa Etária**

A análise da massa de teste permite cruzar os dados demográficos dos clientes com seus hábitos de consumo. A tabela abaixo mostra que o gasto médio por transação (ticket médio) aumenta progressivamente com a idade, sendo mais alto entre os clientes com mais de 51 anos. Este insight é fundamental para direcionar produtos e campanhas de maior valor agregado para o público mais maduro.

```
--> Relação: Gasto Médio por Compra (Ticket Médio) por Faixa Etária
```

faixa_etaria	
0-18	NaN
19-25	NaN
26-35	249.27
36-50	NaN
51-65	307.22
65+	NaN

Name: valor_compra, dtype: float64

- **Relação: Análise Financeira por Tipo de Cupom** A tabela a seguir detalha a performance financeira de cada tipo de cupom. Ela agrega a receita total gerada (`valor_compra`) e o custo total dos cupons (`valor_cupom`), permitindo uma avaliação clara da lucratividade e do retorno sobre o investimento de cada estratégia de desconto.

```
--> Relação: Análise Financeira por Tipo de Cupom
```

	Receita_Total	Custo_Total_Cupons
tipo_cupom		
Cashback	249.27	12.46
Desconto	307.22	77.18

4. Conclusão Final

A primeira etapa do Projeto Interdisciplinar cumpriu com sucesso o objetivo de ENTENDER OS DADOS. O processo de coleta, verificação de qualidade e exploração inicial revelou que, embora as bases de dados brutas apresentem desafios como inconsistências de formato e valores ausentes, o tratamento adequado permite a extração de insights estratégicos valiosos.

A análise exploratória demonstrou a importância de utilizar a base de dados correta para cada tipo de questão de negócio, separando análises de comportamento (base de transações principal) de análises financeiras (massa de teste). Os resultados numéricos iniciais já apontam para uma clara correlação entre o perfil demográfico do cliente (idade) e seu poder de compra, fornecendo uma base sólida para futuras segmentações de mercado e otimização de campanhas.

Este trabalho inicial de tratamento e análise de dados é a fundação essencial sobre a qual o dashboard interativo e as análises mais aprofundadas das próximas entregas serão construídos.