

Case Técnico Data Analytics

André Matos



Objetivo

- Avaliar impacto da campanha de cupons
- Criar segmentações de usuários
- Recomendar próximos passos

Limitações do Estudo

Descrição das tabelas de dados disponíveis

Pedidos (order.json)

<https://data-architect-test-source.s3-sa-east-1.amazonaws.com/order.json.gz>

Contém dados de cerca de 3.6 milhões de pedidos realizados entre dez/18 e jan/19. Cada pedido possui um **order_id** e os seguintes atributos complementares:

- **cpf** (string): Cadastro de Pessoa Física do usuário que realizou o pedido
- **customer_id** (string): Identificador do usuário
- **customer_name** (string): Primeiro nome do usuário
- **delivery_address_city** (string): Cidade de entrega do pedido
- **delivery_address_country** (string): País da entrega
- **delivery_address_district** (string): Bairro da entrega
- **delivery_address_external_id** (string): Identificador do endereço de entrega
- **delivery_address_latitude** (float): Latitude do endereço de entrega
- **delivery_address_longitude** (float): Longitude do endereço de entrega
- **delivery_address_state** (string): Estado da entrega
- **delivery_address_zip_code** (string): CEP da entrega
- **items** (array(json)): Itens que compõem o pedido, bem como informações complementares como preço unitário, quantidade, etc.

- **merchant_id** (string): Identificador do restaurante
- **merchant_latitude** (float): Latitude do restaurante
- **merchant_longitude** (float): Longitude do restaurante
- **merchant_timezone** (string): Fuso horário em que o restaurante está localizado
- **order_created_at** (timestamp): Data e hora em que o pedido foi criado
- **order_id** (string): Identificador do pedido
- **order_scheduled** (bool): Flag indicando se o pedido foi agendado ou não (pedidos agendados são aqueles que o usuário escolheu uma data e hora para a entrega)
- **order_total_amount** (float): Valor total do pedido em Reais
- **origin_platform** (string): Sistema operacional do dispositivo do usuário
- **order_scheduled_date** (timestamp): Data e horário para entrega do pedido agendado

Limitações do Estudo

- O campo `account_id` (32 caracteres) do dataset Pedidos não é compatível com `customer_id` (64 caracteres)

```
# Quantos account_ids estão em df_ab_test?
matched_ids = df['account_id'].isin(df_ab_test['customer_id'])
print(f"IDs que batem: {matched_ids.sum()} de {len(df)} ({(matched_ids.sum()/len(df))*100:.2f}%")
```

✓ 0.2s

IDs que batem: 0 de 306534 (0.00%)

```
# Validação de dados
print("Exemplo do df (transações):", df['account_id'].dropna().unique()[:5])
print("Exemplo do df_ab_test:", df_ab_test['customer_id'].dropna().unique()[:5])
```

✓ 0.4s

Exemplo do df (transações): ['0009655768c64bdeb2e877511632db8f' '00116118485d4dfda04fdbaba9a87b5c'
'0011e0d4e6b944f998e987f904e8c1e5' '0020c2b971eb4e9188eac86d93036a77'
'0020ccbbb6d84e358d3414a3ff76cffd']

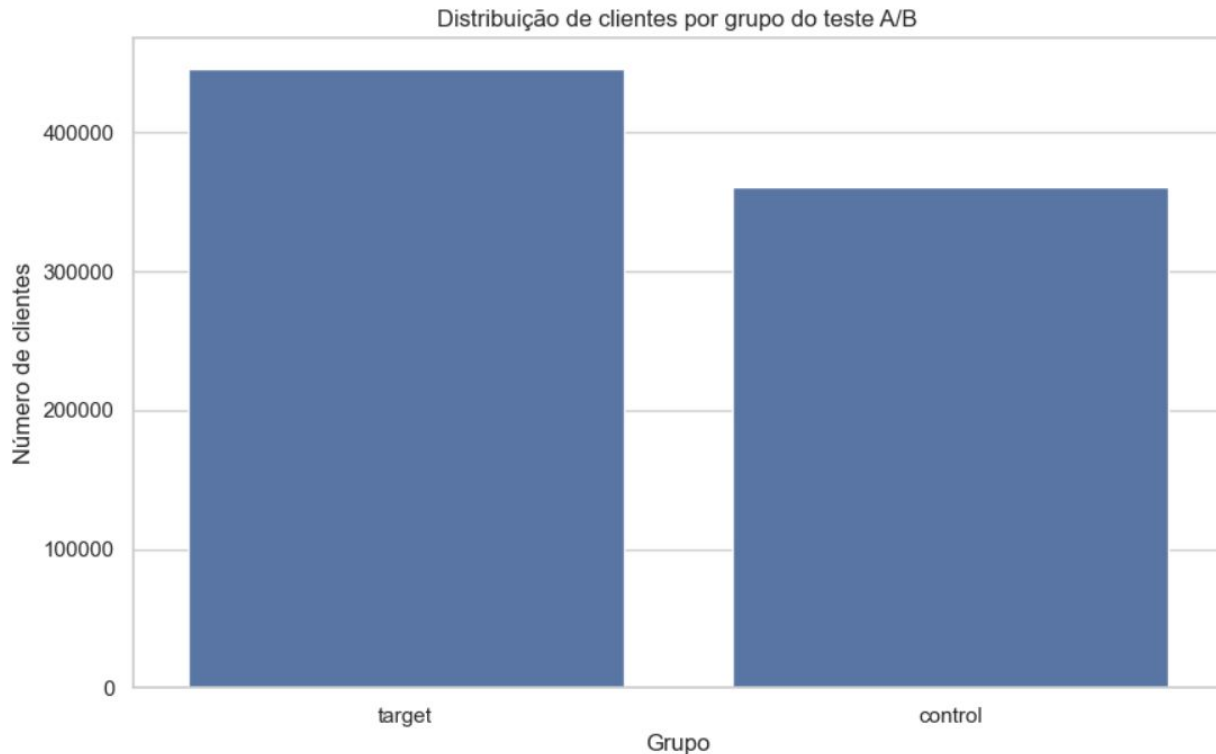
Exemplo do df_ab_test: ['755e1fa18f25caec5edffb188b13fd844b2af8cf5adedcf77c028f36cb9382ea'
'b821aa8372b8e5b82cdc283742757df8c45eecdd72adf411716e710525d4edf1'
'd425d6ee4c9d4e211b71da8fc60bf6c5336b2ea9af9cc007f5297541ec40b63b'
'6a7089eea0a5dc294fbccd4fa24d0d84a90c1cc12e829c8b535718bbc651ab02'
'dad6b7e222bab31c0332b0ccd9fa5dbd147008facd268f5e3763fa657c23a58d']

Limitações do Estudo

- Não foi possível utilizar o dataset Pedidos (transactions.json)
- Análise foi baseada nos dataset Usuários (consumer.csv) e Marcação de usuários que participaram do teste A/B (ab_test_ref.csv)
- Não foi possível medir diretamente:
 - Taxa de conversão
 - Ticket médio
 - Receita gerada
 - Análise de viabilidade financeira da campanha (ROI)

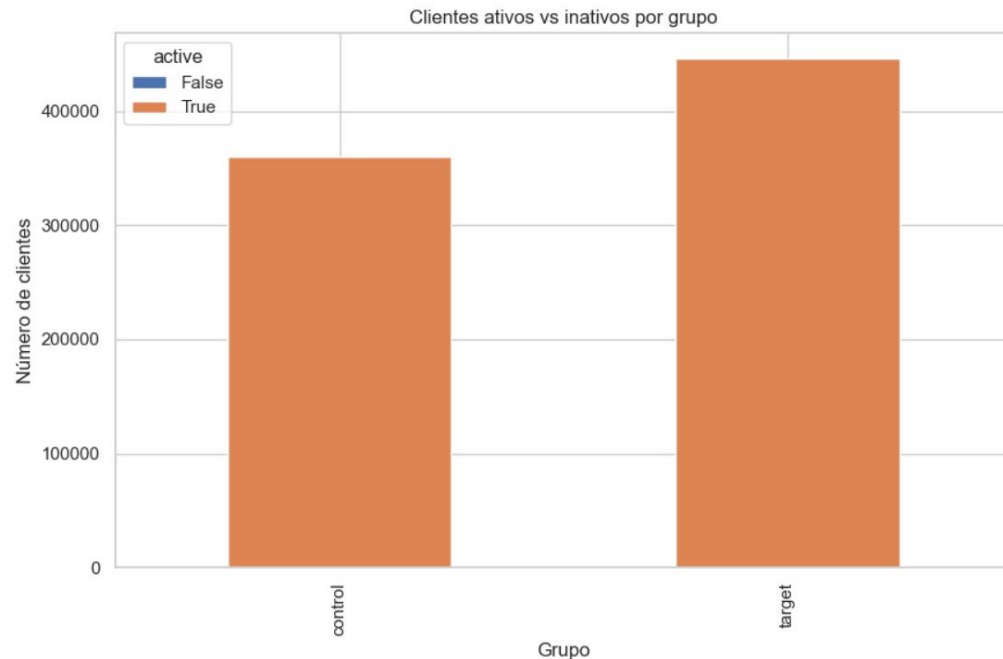
Distribuição de Usuários

- 445,9K target (55,3%) e 360,5K control (44,7%)
- A diferença requer atenção, desbalanceamento pode afetar a significância estatística
- Sem dados de pedidos, não é possível confirmar se o balanceamento se mantém em métricas de conversão ou retenção



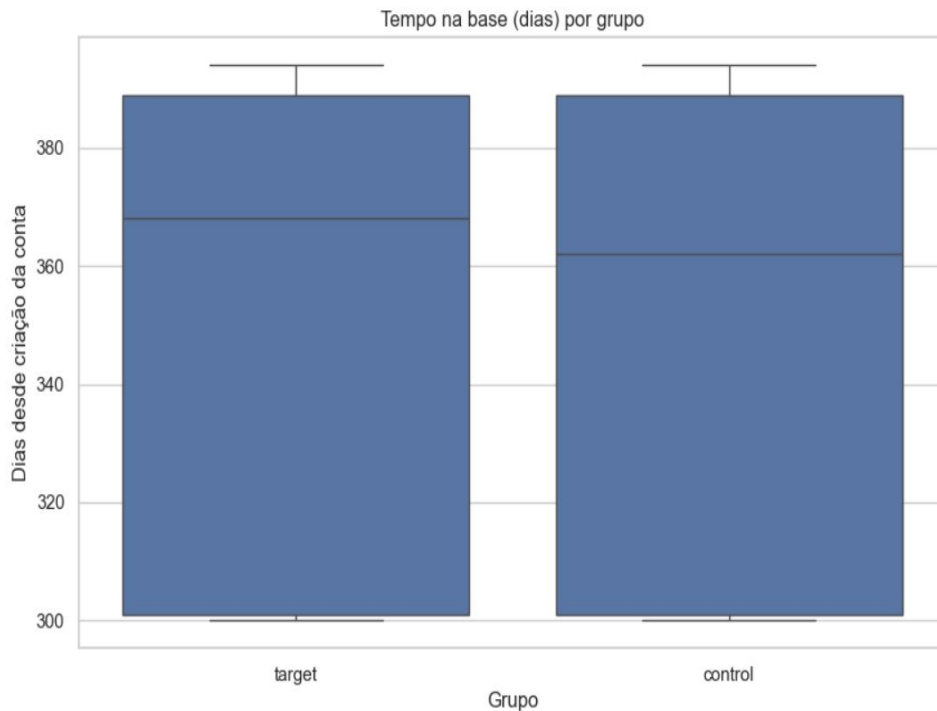
Atividade dos Usuários

- Apenas 882 usuários inativos target e 713 de control (0,2%)
- A estratégia de cupons não impactou significativamente a taxa de inatividade no período analisado
- Se "ativo" inclui qualquer interação (não só compras), a métrica pode estar inflada.



Tempo na Plataforma

- Diferença de apenas 1.6% na mediana (368 vs 362 dias)
- Sugere que outros fatores (como tempo de uso) não influenciaram os resultados
- Cupons podem ter menor impacto (usuários fiéis)
- Sugerir teste com usuários novos (< 90 dias) para avaliar efeito real
- Oferecer benefícios diferenciados para usuários antigos



Conclusão

Apesar da limitação imposta pela ausência dos dados de pedidos/transações, foi possível estruturar e validar um teste A/B com boa qualidade técnica. As análises demonstraram que:

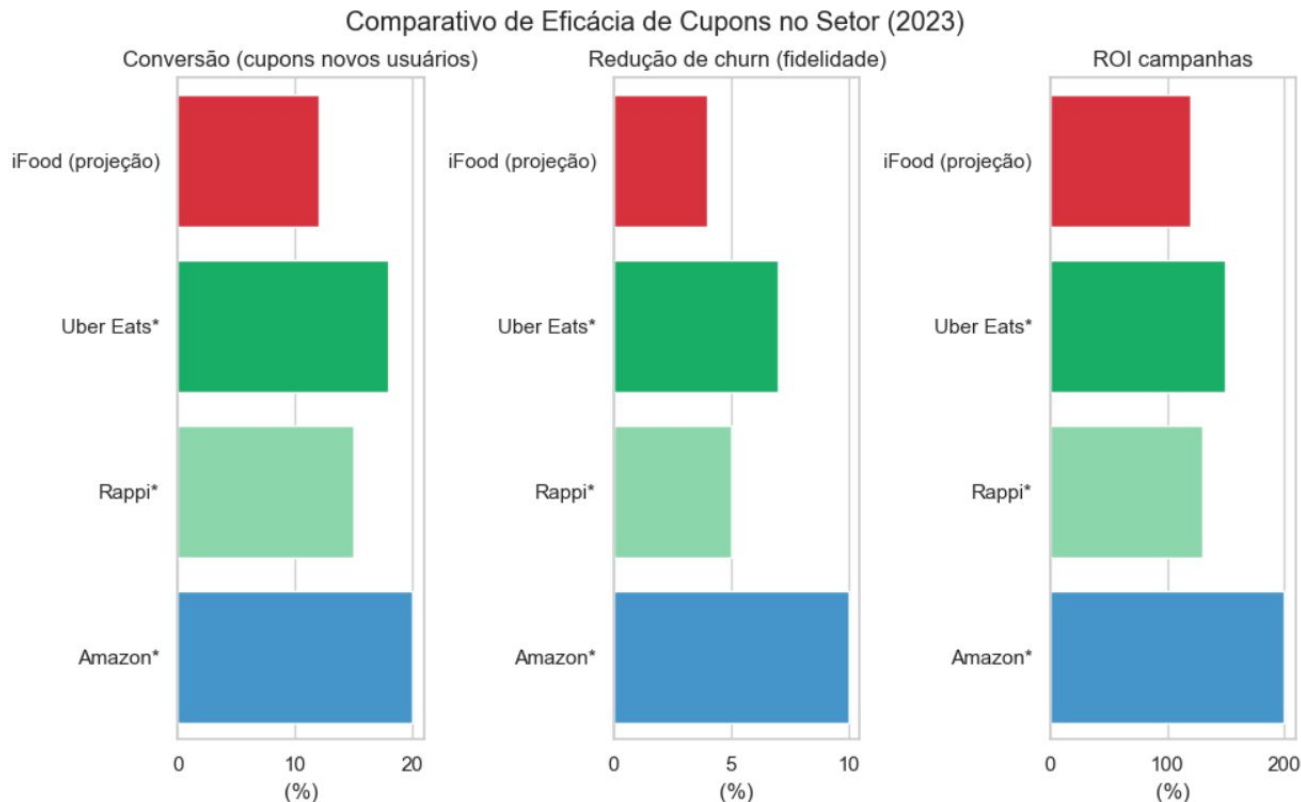
- A proporção 55-45% são geralmente aceitas como balanceadas, desde que a randomização seja válida.
- Taxa de inatividade insignificante (0.2% em ambos grupos), porém sem dados de compras para validar o engajamento
- Usuários majoritariamente antigos (mediana de ~1 ano na plataforma), o que pode limitar o impacto de cupons

Sugestões

- Usuários novos (< 90 dias): Testar cupons de boas vindas (ex: 20% OFF no 1º pedido)
- Usuários antigos (> 1 ano): Oferecer benefícios de fidelidade (ex: frete grátis após x pedidos)
- Redesenhar testes A/B com grupos perfeitamente balanceados (50-50%)
- Critérios claros de “ativo” (ex: ≥ 1 pedido nos últimos 30 dias)

Sugestões

- Sem os dados de transações, não podemos afirmar com exatidão o ROI da campanha. No entanto, experiências de outras plataformas sugerem que estratégias segmentadas podem aumentar a conversão em 10-15% e redução de churn em 3-5%.



Fonte: Dados de relatórios públicos de 2023

Obrigado!

