

Case Técnico Data Science - iFood

Visão Geral

Neste case técnico, você deverá desenvolver uma solução baseada em dados para otimizar a distribuição de cupons e ofertas aos clientes do iFood. O desafio permitirá que você demonstre suas habilidades em análise de dados, modelagem e comunicação de resultados.

Contexto

Para aumentar o engajamento dos clientes, as empresas fazem uso de cupons que são enviados por meio de vários canais de marketing. Essa estratégia não apenas atrai os clientes para um serviço, mas também pode estabelecer um relacionamento de longo prazo, aumentando a probabilidade de transformar clientes ocasionais em consumidores recorrentes.

No entanto, ter uma estratégia eficiente de distribuição de ofertas é um desafio considerável devido aos múltiplos aspectos envolvidos:

- Diferentes tipos de ofertas disponíveis
- Diversos canais de marketing
- Variados perfis de consumo dos clientes

- Timing do envio das ofertas

Objetivo

Você deverá:

1. Analisar os dados históricos de transações, ofertas e clientes
2. Desenvolver uma técnica/modelo que auxilie na decisão de qual oferta enviar para cada cliente
3. Demonstrar o potencial impacto da sua solução no negócio

Dados Disponíveis

O projeto utiliza três conjuntos de dados:

offers.json

Contém os ids das ofertas e metadados de cada uma delas:

- id (string): id da oferta
- offer_type (string): o tipo da oferta (BOGO, discount, informational)
- min_value (int): valor mínimo para ativação da oferta
- duration (int): duração da oferta
- discount_value (int): valor do desconto
- channels (list of strings): canais de veiculação

customers.json

Contém atributos de cerca de 17k clientes:

- age (int): idade do cliente na criação da conta
- registeredon (int): data de criação da conta
- gender (string): gênero do cliente
- id (string): id do cliente
- credit_card_limit (float): limite do cartão registrado

transactions.json

Contém cerca de 300k eventos:

- event (str): descrição do evento (transação, oferta recebida, etc.)
- account_id (str): id do cliente
- time_since_test_start (int): tempo desde o começo do teste em dias (t=0)
- value (json): registra offer_id, desconto (reward) ou valor da transação

Os dados podem ser acessados através desse link:

<https://data-architect-test-source.s3.sa-east-1.amazonaws.com/ds-technical-evaluation-data.tar.gz>

O Desafio

Você deverá entregar:

1. **Notebook de Processamento de Dados**

2. Manipulação e limpeza dos dados
3. Preparação de dataset unificado
4. Deve utilizar PySpark
5. Recomendamos usar Databricks Community Edition (<https://community.cloud.databricks.com/>)
6. **Notebook/Script de Modelagem**
7. Processo de treino e avaliação do modelo escolhido
8. Use as ferramentas/linguagens com as quais se sentir mais confortável
9. A escolha do modelo e métricas ficam a seu critério
10. **Apresentação (máx. 5 slides)**
11. Destinada a líderes de negócio tecnicamente leigos
12. Deve conter estimativas ou projeções de performance da estratégia proposta

Estrutura do Repositório

```
ifood-case/
├── data/                # Datasets
│   ├── raw/            # Dados originais
│   └── processed/      # Dados processados
├── notebooks/          # Jupyter notebooks
│   ├── 1_data_processing.ipynb
│   └── 2_modeling.ipynb
├── presentation/        # Slides para stakeholders
├── src/                 # Código fonte (opcional)
├── README.md
└── requirements.txt
```

Critérios de Avaliação

Serão avaliados:

- Qualidade e organização do código
- Processo de análise exploratória
- Justificativa das escolhas técnicas
- Criatividade na solução proposta
- Clareza na comunicação dos resultados

Instruções de Entrega

1. Crie um repositório público ou privado no GitHub
2. Desenvolva sua solução
3. Atualize o README com instruções de execução
4. Envie o link do seu repositório

Dúvidas

Em caso de dúvidas sobre o desafio, abra uma issue neste repositório.

Boa sorte!