

Desafio para Estágio de Dados

Descrição do desafio

Na Sallve, nós co-criamos. Acreditamos muito na importância de ouvir a nossa comunidade, pois é assim que vamos continuar a desenvolver "fórmulas incríveis inspiradas por necessidades reais". Pensando nisso, produzimos o nosso quiz da pele, que possui centenas de milhares de respostas, e nos dá insumos para entender as reais necessidades da nossa comunidade.

Você vai receber, junto a essa descrição do desafio, uma amostra anonimizada do nosso quiz (*quiz-sample.csv*), um conjunto também anonimizado de dados de compras de alguns dos nossos produtos (*orders-sample.csv*), alguns dados sobre o investimento de mídia feito na promoção de cada um desses produtos no mesmo período dos dados de compras (*product-marketing-investment.csv*) e um recorte dos últimos 30 dias da nossa base de facebook (*last-30d-facebook.csv*).

Com esses dados, esperamos que você consiga responder às seguintes perguntas:

A) Considerando a base de Facebook Ads, responda:

1. Como você realocaria o investimento entre as campanhas de Facebook Ads a fim de otimizar o seu retorno sobre investimento?
2. Qual é a(s) principal(is) métrica(s) que você acredita que exemplifica o potencial de conversão de um anúncio? Baseado na sua resposta, quais campanhas pausaria?
3. Dentre as campanhas apresentadas, qual delas apresenta o maior potencial para escalar o investimento?

B) Considerando as bases de quiz e compras, responda:

1. Quais as 5 preocupações de pele mais recorrentes entre os nossos clientes a cada mês, considerando apenas respostas de pessoas que possuam pedidos registrados na base de compras?
2. Observando como as preocupações e seus pesos relativos se movimentam ao longo do período analisado na resposta do item 1:
 - quais são as hipóteses que poderiam justificar esses movimentos?
 - como você validaria essas hipóteses?

Vale frisar que, para a análise do item 2, queremos entender como você aborda o problema, explora os dados, constrói e valida (ou tentaria validar) as suas hipóteses, e como apresenta as suas conclusões.

A seguir, você encontra as descrições dos campos, seus formatos e tipos para cada um dos conjuntos de dados disponibilizados.

Formato dos dados para a parte B

Quiz da pele

Para cada pessoa, conta com múltiplas linhas, cada linha associada a uma preocupação de pele. Contém os seguintes campos:

- **person_id:** string, identificador anonimizado do participante, correspondente à coluna de mesmo nome no dataset de compras
- **genero:** string, gênero com o qual a pessoa se identifica
- **idade:** string, intervalo de idade da pessoa
- **tipo_pele:** string, descreve o tipo de pele da pessoa
- **pele_sensivel:** inteiro, binário, indica se a pessoa se descreve como tendo pele sensível
- **estado_vive:** string, estado onde a pessoa vive
- **preocupacao_de_pele:** string, nome atribuído a uma preocupação de pele, possui valores no formato pp_[nome-da-preocupação], que indicam possíveis preocupações e pp_nao_tenho_preocupacoes que é a opção que indica que a pessoa não se identifica com nenhuma das preocupações possíveis.
- **tem_preocupacao:** inteiro, binário, indica se a pessoa se descreve como tendo a preocupação de pele da linha correspondente.
- **data-submissao:** data na qual a pessoa enviou as respostas.

Dados de compra:

Cada linha corresponde a um pedido de compra recebido. Contém os seguintes campos:

- **person_id:** string, identificador anonimizado do comprador, correspondente à coluna de mesmo nome no dataset do quiz
- **product_[A-E]:** inteiro, quantidade do produto em questão incluída no pedido
- **processed_at:** datetime, horário de processamento do pedido
- **cancelled_at:** datetime, se o pedido foi cancelado, contém horário do cancelamento
- **financial_status:** string, status de pagamento do pedido
- **billing_city:** string, cidade do endereço de cobrança
- **billing_state:** string, estado do endereço de cobrança

Dados de investimento de mídia

Cada linha corresponde ao investimento feito no produto e mês correspondentes. Contém os seguintes campos:

- **report_month:** mês do investimento
- **product:** nome do produto, no mesmo formato dos dados de compra

- **investment:** volume de investimento de mídia em reais (R\$)

Formato de entrega

Para nos enviar a sua análise, você pode optar por um formato popular de documento (PDF, slides, .doc/.docx, etc) ou Notebooks (.ipynb), acompanhado de qualquer código que tenha produzido (SQL, Python, R, Scala, etc), junto às instruções de execução.

Caso tenha qualquer dúvida, não hesite em nos contatar.