

Avaliação de Cientista de Dados

Bem-vindo(a) a avaliação de Cientista de Dados da Hotmart. Nesta avaliação, iremos testar seus conhecimentos em campos como análise, desenvolvimento e estatísticas. Para tal, pedimos que se atente às instruções abaixo sobre como realizar sua prova. Vamos lá?

Contexto

Na Hotmart, possuímos três principais *personas* que integram nosso negócio: os produtores, os afiliados e os compradores.

- **Produtores são pessoas que criam produtos digitais na Hotmart**, como cursos de idiomas, ebooks de receitas culinárias, audiolivros, softwares, dentre muitos outros exemplos.
- **Afiliados são pessoas que promovem produtos dos produtores** em troca de uma comissão na venda, que varia de produto para produto, e de afiliado para afiliado.
- **Compradores são pessoas que adquirem um ou mais produtos digitais.**

Uma venda é feita por um afiliado quando alguém clica em um link de afiliados.

Eles geralmente fazem a promoção desses produtos em redes sociais, vídeos, anúncios, etc.

Já uma venda é feita por um produtor quando alguém tem acesso direto ao seu produto, sem intermediação do afiliado. Por exemplo, pessoas que seguem o Whindersson Nunes no Youtube e entraram em seu site oficial para adquirir seu produto, ou clicaram no link do produto sem código de afiliação.

Coleta de dados

Utilizando o [MySQL Workbench](#) (ou outro cliente SQL de sua preferência), conecte-se ao endereço abaixo:

- Host:** interview-2.ck1h5ksgzpiq.us-east-1.rds.amazonaws.com
- Porta:** 3306
- Usuário:** hotinterview
- Senha:** 6cT4jk9QWPhQC9KXWKDd
- Schema:** innodb
- Tabela:** sales_data

Sobre os dados

Durante sua avaliação, você irá analisar dados referentes a uma amostra de **compras feitas na Hotmart em 2016**. Tratam-se de mais de **1,5 milhão registros** de compras realizadas em nossa plataforma. Abaixo, nós iremos detalhar o que significa cada campo:

- **purchase_id**: Identificação da compra na Hotmart;
- **product_id**: Identificação do produto na Hotmart;
- **affiliate_id**: Identificação do afiliado na Hotmart;
- **producer_id**: Identificação do produtor na Hotmart;
- **buyer_id**: Identificação do comprador na Hotmart;
- **purchase_date**: Data e hora em que a compra foi realizada;
- **product_creation_date**: Data e hora em que o produto foi criado na Hotmart;
- **product_category**: categoria do produto na Hotmart. Exemplo: e-book, software, curso online, e-tickets, etc.;
- **product_niche**: nicho de mercado que o produto faz parte. Exemplo: educação, saúde e bem-estar, sexualidade, etc.;
- **purchase_value**: valor da compra. Esse dado, assim como nicho e categoria foi codificado para manter a confidencialidade. O valor apresentado no dataset é o z-score do valor real;
- **affiliate_commission_percentual**: percentual de comissão que o afiliado receberá da compra;
- **purchase_device**: tipo de dispositivo utilizado no momento da compra, como: Desktop, Mobile, Tablet, ou Outros;
- **purchase_origin**: endereço do site do qual a pessoa veio antes da compra. Por exemplo, se uma pessoa veio do Facebook, Youtube, ou até mesmo de outra página no site oficial do produto;
- **is_origin_page_social_network**: informa se essa compra veio de uma URL do Facebook, Youtube, Instagram, Pinterest, ou Twitter.

Algumas regras de negócio

- Quando a compra for feita diretamente pelo produtor, ou seja, quando não houver afiliado intermediando a compra, o campo **affiliate_commission_percentual** terá valor **0**, e o campo **affiliate_id** será igual ao **producer_id**;
- No campo **purchase_origin** nós apenas consideramos o host do site. Isso quer dizer que, se uma pessoa veio do site **www.meuproduto.com/promocoes**, esse campo só irá retornar o valor **www.meuproduto.com**;

Sobre a avaliação

Nessa avaliação, nós queremos que você mostre sua linha de pensamento durante a condução de uma pesquisa/análise. Essa avaliação representa muito o dia a dia de um Cientista de Dados na Hotmart, por isso, é muito importante sabermos como você responderia a perguntas de negócio.

Sendo assim, **sua tarefa será gerar Insights sobre os dados no dataset** que disponibilizamos. Não existe uma forma única de chegar a uma solução, nem um número mínimo ou máximo de insights para serem gerados, mas, separamos algumas perguntas que seriam interessantes de serem respondidas:

- A Hotmart depende dos maiores produtores da plataforma? Ou seja, os produtores que mais vendem são responsáveis pela maior parte do faturamento da Hotmart?
- Existe algum padrão ou tendência relevante nos dados?
- É possível segmentar os usuários com base em suas características (faturamento, nicho de produto, etc.)?
- Quais características mais impactam no sucesso de um produto? Ou seja, o que faz um produto vender mais?
- É possível estimar quanto de faturamento a Hotmart irá fazer nos próximos três meses a partir do último mês mostrado no dataset?

Nessa prova nós iremos avaliar os seguintes pontos na sua solução. Vale lembrar que mais importante que sua solução é o caminho que você levou até chegar a ela. **Não há problema em mostrar soluções que não deram certo durante sua avaliação.**

- Storytelling + Visualização de Dados
- Análise Exploratória de Dados (EDA)
- Melhores práticas em desenvolvimento de software
- Análise preditiva
- Estatísticas

Entrega da avaliação

Após realizar as etapas da avaliação, envie os arquivos que referentes a ela para o email: rafaela.euzebio@hotmart.com