

Passei Direto

Teste - Vaga Data Scientist

O objetivo desse teste é avaliar como você se comporta diante de um conjunto de dados e questões que queremos responder. Fique a vontade para ir além do perguntado e trazer insights extras!

Os dados para a análise estão disponíveis em anexo. Eles correspondem a uma amostra dos dados gerados pela plataforma do Passei Direto e são melhor detalhados a seguir, junto com as questões do teste.

1ª Parte

O Passei Direto é uma plataforma aberta (grátis) com um modelo de negócios freemium, baseado em assinaturas. É muito importante para a saúde do negócio entendermos bem o comportamento dos usuários, principalmente dos pagantes.

Nessa etapa, será usada a *BASE A*. Essa base consiste de apenas um Dataset:

- **premium_students.json** - Registros de usuários cadastrados no Passei Direto em Novembro de 2017 e que também assinaram o Plano Premium

Questão 1: Dentre os usuários cadastrados em Nov/2017 que assinaram o Plano Premium, qual a probabilidade do usuário virar Premium após o cadastro em ranges de dias? A escolha dos ranges deve ser feita por você, tendo em vista os insights que podemos tirar para o negócio.

Exemplo de distribuição:

- De 0 a 3 dias -> 30%
- De 4 a 8 dias -> 20%
- De 8 a 30 dias -> 50%

2ª Parte

Nessa segunda etapa, vamos fazer uma análise mais completa dos nossos usuários e suas ações no Passei Direto. Será usada a *BASE B*, composta dos seguintes Datasets:

- **students.json** - Amostra de usuários que acessaram o Passei Direto no mês de Novembro de 2017. Todos os outros datasets referenciam esses usuários
- **sessions.json** - Todas as sessões dos usuários ao longo do tempo
- **subjects.json** - Disciplinas seguidas pelos usuários

- **questions.json** - Perguntas feitas pelos usuários
- **answers.json** - Respostas à perguntas feitas pelos usuários
- **fileViews.json** - Arquivos visualizados pelos usuários
- **studyPlanViews.json** - Visualizações dos conteúdos de Planos de Estudo pelos usuários. Aqui são incluídos tanto os vídeos quanto os resumos
- **textBookSolutionViews.json** - Visualização de Exercícios Resolvidos dos livros pelos usuários
- **evaluations.json** - Avaliações dos conteúdos feitas pelos usuários. São destacadas tanto as avaliações positivas (Likes) quanto as negativas (Dislikes)
- **premium_payments.json** - Pagamentos realizados pelos usuários que assinaram o Plano Premium do Passei Direto
- **premium_cancellations.json** - Pedidos de cancelamento do Plano Premium solicitados pelos usuários pagantes

Questão 2: Faça uma análise geral dos dados e apresente as informações que julgar mais relevantes dessa base.

Questão 3: Faça uma análise comparativa do comportamento dos usuários não Premium e dos usuários Premium. Que tipos de ações podemos direcionar para usuários não Premium fazerem com o objetivo de termos um maior número de assinantes?

Questão 4: Em Novembro de 2017 fizemos uma grande mudança no PD: o Content Restriction. Os usuários não Premium passaram a poder consumir no máximo 3 arquivos diferentes por mês. Diante dessa mudança, qual passou a ser o Lifetime Value (LTV) dos usuários Premium a partir de Novembro de 2017?

Questão 5: Entre os usuários que "Churnaram" solicitando ativamente o cancelamento do Plano Premium, o que está fortemente correlacionado com o cancelamento? (Ex: uso das features, tempo de inatividade, etc)

Questão 6: Quais são as 5 maiores universidades no Passei Direto? E quais são os principais tipos de Arquivos consumidos em cada uma delas?

Questão 7: Considerando o seu conhecimento da plataforma do Passei Direto e a análise dos dados disponibilizados nesse teste, que oportunidades você vê para melhorarmos a experiência dos nossos usuários? Ofereça insights explorando o como e o porquê.

Entregáveis

- Código-fonte utilizado para fazer o processamento e análise dos dados
 - Idealmente colocar em um repositório no Github
 - Enviar por email até o final do prazo combinado

- Apresentação dos resultados
 - Vamos marcar um dia para você apresentar presencialmente
 - Formato de sua livre escolha
 - Importante explicar também o seu processo de decisão diante das questões apresentadas
 - Levar em conta que as pessoas para quem você vai apresentar podem não ter necessariamente um background técnico da área de dados

Boa sorte, e nos vemos em breve.