

OppaiMan - Desafio Analista de Dados Sênior

Este desafio foi elaborado para avaliar conhecimentos essenciais para a posição de Analista de Dados Sênior na OppaiMan. Serão avaliados aspectos relacionados à análise de dados, modelagem de dados, comunicação de resultados e muito mais.

É esperado que você resolva os problemas propostos de forma eficiente e criativa, demonstrando seu domínio em técnicas analíticas e habilidades de resolução de problemas complexos.

Esperamos também que você se divirta e, quem sabe, aprenda algo no processo. 😊

Instruções

Este desafio é bem diferente dos testes tradicionais realizados por outras empresas. Nele, nós reconhecemos o valor do seu tempo investido para a realização do teste e, como forma de incentivo, recompensamos financeiramente por todas as horas dedicadas. Nosso objetivo é motivar os candidatos a entregar os melhores resultados possíveis.

Leia atentamente a descrição do desafio. Você terá um prazo máximo de 2 semanas para concluir a tarefa. Após a leitura, estime a quantidade de horas que planeja investir no teste e compartilhe essa estimativa com o avaliador. Além disso, será necessário informar a data de início do teste, bem como fornecer a chave PIX para receber o pagamento correspondente à primeira metade equivalente às horas estimadas. Queremos proporcionar a você uma experiência tranquila e valorizar seu esforço desde o início.

Obs¹.: A remuneração por hora é de R\$ 150,00.

Obs².: Embora seja um teste remunerado, os arquivos criados durante o desafio pertencem exclusivamente ao candidato, preservando todos os seus direitos sobre eles.

Obs³.: Para cumprir com as obrigações fiscais, solicitaremos a emissão de uma Nota Fiscal (NF) para os candidatos que possuem CNPJ. Caso o candidato não possua CNPJ, emitiremos um Recibo de Pagamento Autônomo (RPA) para garantir a conformidade com as regulamentações fiscais. Queremos manter nossa contabilidade atualizada e garantir uma experiência transparente para todos os envolvidos.

Contexto Técnico

Na Oppai-Man, oferecemos uma variedade de serviços interconectados, todos projetados para criar uma plataforma onde nossos clientes podem

interagir com nossa comunidade, compartilhar sugestões para aprimorar nossos jogos e adquirir episódios inéditos de nossos títulos. Cada um desses serviços requer dados específicos para operar de maneira eficaz.

Para auxiliar nossas equipes na tomada de decisões, dispomos de uma Plataforma de Dados que viabiliza a ingestão, processamento e representação visual de informações.

Uma das componentes fundamentais de nossa plataforma é o repositório de dados, conhecido como Data Lake.

Descrição do problema

Uma parte essencial da plataforma Oppai-Man é a interação entre os usuários por meio de postagens e enquetes, juntamente com a venda de episódios através de subscrições do Patreon. Para alcançar isso, armazenamos essas informações em várias coleções do MongoDB, seguindo o seguinte modelo de estrutura:

Coleção "usuários":

- `_id`: Identificador único para cada usuário.
- `created_at`: Data de criação do usuário.
- `country`: País do usuário.
- `language`: Idioma preferido do usuário.
- `updated_at`: Data de atualização do perfil do usuário.
- `patreon_status.is_patron`: Indica se o usuário é um apoiador (sim/não).
- `patreon_status.updated_at`: Data de atualização do status do Patreon do usuário.
- `patreon_status.campaign_lifetime_support_cents`: O valor total em centavos que o usuário apoiou ao longo da vida da campanha do Patreon.
- `patreon_status.currently_entitled_amount_cents`: O valor atualmente devido pelo usuário em centavos.
- `patreon_status.is_follower`: Indica se o usuário é um seguidor do Patreon (sim/não).
- `patreon_status.last_charge_date`: Data da última cobrança do Patreon do usuário.
- `patreon_status.last_charge_status`: Status da última cobrança do Patreon do usuário.
- `patreon_status.lifetime_support_cents`: O valor total em centavos que o usuário apoiou ao longo da vida no Patreon.
- `patreon_status.pledge_relationship_start`: Data de início do relacionamento de apoio no Patreon.
- `patreon_status.will_pay_amount_cents`: O valor que o usuário se compromete a pagar em centavos.
- `patreon_status.is_free_trial`: Indica se o usuário está em um período de teste gratuito (sim/não).
- `timezone.name`: Nome do fuso horário do usuário.
- `timezone.offset`: Deslocamento de fuso horário do usuário em relação ao UTC.

- `timezone.is_dst`: Indica se o horário de verão está ativo para o fuso horário do usuário (sim/não).

Coleção "postagens":

- `_id`: Identificador único para cada postagem.
- `type`: Tipo de postagem, seja "enquete" ou "postagem simples".
- `created_at`: Data de criação da postagem.
- `updated_at`: Data de atualização da postagem.
- `status`: Status da postagem.
- `translations`: Um objeto contendo traduções da postagem para diferentes idiomas, incluindo:
 - `title`: Título da postagem.
 - `description`: Descrição da postagem.
 - `options`: Opções para enquetes (se aplicável).
 - `language`: Idioma da tradução.
 - `created_at`: Data de criação da tradução.
 - `updated_at`: Data de atualização da tradução.
 - `multiple_votes`: Se a postagem for enquete, poderá ser de voto múltiplo ou não;

Coleção "comentários de postagens":

- `_id`: Identificador único para cada comentário de postagem.
- `user_id`: Identificação do usuário que deixou o comentário.
- `body`: Conteúdo do comentário.
- `created_at`: Data de criação do comentário.
- `updated_at`: Data de atualização do comentário.
- `post_id`: Identificação da postagem à qual o comentário está associado.

Coleção "votos":

- `_id`: Identificador único para cada voto.
- `user_id`: Identificação do usuário que votou.
- `votes`: Um array de votos múltiplos/simples e com/sem peso de votação;
- `created_at`: Data de criação do voto.
- `updated_at`: Data de atualização do voto.
- `post_id`: Identificação da postagem para a qual o voto foi expresso.

Essas coleções de dados estão relacionadas e podem ser usadas para gerenciar postagens, comentários, votos em enquetes e informações de usuários. Também, coletamos informações valiosas dos usuários que são patronos (Pessoas que financiam o projeto).

Parte 1

Realize uma análise exploratória dos dados fornecidos em anexo a este documento. É fundamental documentar todo o processo e apresentar as conclusões de forma clara e eficaz ao término da análise.

Parte 2

Na segunda etapa, você deverá criar uma modelagem de dados que funcione de forma incremental para os dados fornecidos, garantindo que ela facilite a análise de dados e a geração de relatórios. A modelagem deve ser projetada de tal forma que, se novos dados forem adicionados nas coleções, eles se encaixarão perfeitamente nesse novo modelo.

Após a modelagem, responda as seguintes perguntas:

- Porque você modelou dessa forma?
- Essa modelagem serviria pra um treinamento de um modelo de machine learning?
- As estruturas das collections fornecidas podem ser melhoradas para uma melhor análise dos dados? Se sim, sugira melhorias para as estruturas.

Obs.: Disponibilizar na nuvem e oferecer recursos de visualização de dados são opcionais, mas agregam valor na avaliação do desafio;

Resumo do que é esperado

- Mostrar resultados da análise exploratória dos dados fornecidos;
- Criar uma modelagem para os dados, justificando as escolhas e hipóteses;
- Implementar um processo de ETL que transforme os dados das collections do db na nova modelagem proposta;
- Propor e justificar métricas relevantes para o contexto;

Considerações finais

Os arquivos referentes aos dados estão anexadas junto a esse documento. Os arquivos deverão ser entregues por meio de repositórios git. Em caso de dúvidas relacionadas ao desafio, contate o avaliador através do Discord ou LinkedIn.

Boa sorte!