

Parabéns por ter chegado até esta etapa do processo seletivo! Nesta etapa teremos uma simulação de um possível caso relacionado ao dia a dia do trabalho na empresa. É só seguir as instruções abaixo no cenário proposto. Boa sorte no desafio!

Escopo do Desafio

O novo aplicativo de streaming "5GFlix" está com o desafio de fazer estudos de mercado para formular a estratégia de negócio que irão adotar.

Para uma parte do estudo de mercado, a "5GFlix" precisa fazer análises em cima de filmes e séries que estão disponíveis na Netflix, a sua concorrente direta.

O CTO da "5GFlix", Alan Turing, entrou em contato com a Solvimm para construir uma estrutura lógica que possibilite que o time de BI da "5GFlix" responda a várias perguntas de negócio relacionadas aos dados da Amazon e da Netflix, detalhadas nos entregáveis deste desafio.

Para poder realizar as análises foi fornecida a seguinte base de dados:

- Base1:

<https://drive.google.com/file/d/1gLsCjaMrL91ECdThq58cZAzB9tPxG18g/view?usp=sharing>

- Base2:

https://drive.google.com/file/d/1C_T1w8fc7Oa8MeTo4LMTEcv90IfEOS-6/view?usp=sharing

Para solucionar o desafio, precisamos definir a estrutura para o processamento de dados e também o código de acordo com a necessidade do cliente. O código de todas as funções deve estar documentado e suas entradas e saídas devem ser claras e explicadas na documentação.

Você pode enviar dúvidas sobre o escopo de desafio para o email -----@gmail.com dentro do prazo para dúvidas.

Descrição do Dataset:

Base 1:

1. ID do filme
2. título e ano de lançamento

Base 2:

1. Cust_Id: ID do customer que fez a avaliação
2. Rating: avaliação (nota)

3. Date: data da avaliação
4. Movie_Id: ID do filme

ENTREGÁVEIS

1. Jupyter Notebook ou arquivo Python que responda as seguintes perguntas:
 - 1.1. Quantos filmes estão disponíveis no dataset?
 - 1.2. Qual é o nome dos 5 filmes com melhor média de avaliação?
 - 1.3. Quais os 5 anos com menos lançamentos de filmes?
 - 1.4. Quantos filmes que possuem avaliação maior ou igual a 4.7, considerando apenas os filmes avaliados na última data de avaliação do dataset?
 - 1.5. Dos filmes encontrados na questão anterior, quais são os 10 filmes com as piores notas e quais as notas?
 - 1.6. Quais os id's dos 5 customer que mais avaliaram filmes e quantas avaliações cada um fez?
2. Documentação
 - 2.1. Explicação do passo a passo utilizado no código
 - 2.2. Explicação de como executar o código do zero
3. Você deverá compartilhar seu código em formato Zip ou enviar o repositório Git para que o gestor da posição, acompanhado da documentação entregue no item 2, consiga testar sua solução.

OBS:

- O desenvolvimento pode ser feito localmente no seu computador pessoal, no entanto, recomendamos o uso de alguns dos notebooks gratuitos que existem na internet: ○
 - Google Colaboratory: <https://colab.research.google.com/>
 - Kaggle Notebooks: <https://www.kaggle.com/>

Os entregáveis devem ser enviados em um único e-mail para -----@gmail.com com o assunto "[Nome do Candidato] Desafio Estágio em Dados".

PRAZOS

- Dúvidas sobre o desafio: 4 dias
- Entrega do desafio: 7 dias

TECNOLOGIAS

- Python (pandas)

- Jupyter Notebook