

## TESTE - AVALIAÇÃO PARA A POSIÇÃO DE ENGENHEIRO DE DADOS JR.

O prazo final para entrega da solução é 28/03/2022 às 21h (dois dias úteis), com ou sem todas as soluções.

- Após esse prazo não serão aceitas entregas.
- Antes desse prazo, entregas com todas as respostas terão um adicional na nota final.

### Problema 1 [Extração e Transformação]:

A ANVISA - Agência Nacional de Vigilância Sanitária - publica regularmente, no Diário Oficial da União, dados sobre produtos saneantes; um cliente gostaria que fossem capturadas as publicações realizadas sobre esse tema nos dois últimos meses - janeiro e fevereiro de 2022 - e recebê-las formatadas em uma planilha Excel (layout em anexo).

Orientação:

Link para o diário oficial: <https://www.in.gov.br/leiturajornal> Termo de pesquisa: “deferir os registros e as petições dos produtos saneantes”.

Exemplo de página com dados a serem capturados: [link de exemplo para uma publicação do diário oficial](#)

#### Solução esperada:

##### *Captura*

Você deve utilizar uma forma automática de capturar os dados. Esse passo pode ser feito usando alguma ferramenta de *webscraping* / *crawler* ou através de código (preferencialmente *python*).

Entrega esperada: Descrição de como esse passo foi executado, citando as ferramentas/linguagens utilizadas. Informação de quantas publicações foram capturadas, tempo total de duração do processo etc. Formato livre.

##### *Transformação*

- 1.1. Você deve organizar os dados capturados no layout de planilha solicitado pelo cliente. (anexo)
- 1.2. Você sugere algum tipo de tratamento de transformação e/ou qualidade para os dados extraídos? Quais?

Entrega esperada: planilha Excel com os dados obtidos. // sugestões, se houver

### Problema 2 [Carga e Processo]:

Observando a oportunidade de desenvolver uma análise de dados históricos de saneantes (resultante da solução do Problema 1) a Move decidiu criar um modelo de dados para armazenar tais informações e um processo recorrente de captura e armazenamento.

#### Solução esperada:

##### *Modelagem*

- 1.1. Você deve criar um modelo de dados contendo uma tabela fato para Saneantes e duas tabelas dimensão: uma para Assunto e outra para Categoria.
- 1.2. Você sugere alguma outra tabela dimensão?

Entrega esperada: desenho do modelo com as tabelas, suas colunas, seus relacionamentos e suas chaves. Formato livre.

##### *Processo*

- 1.1. Você deve desenhar um processo de captura, transformação e carga que explique quais passos serão seguidos todos os meses para incrementar novos dados ao modelo.

Entrega esperada: desenho do processo, explicando o que deve ser feito e qual ferramenta será usada em cada passo. Formato livre.

**Entrega final:** as respostas produzidas devem ser enviadas para [marcio.vieira@movecons.com](mailto:marcio.vieira@movecons.com) . O título do e-mail deve conter: **[EDJR] – seu nome**

ANEXO – Layout para a planilha Excel, com uma linha de exemplo.

RESOLUCAO	EMPRESA	AUTORIZACAO	MARCA	PROCESSO	REGISTRO	VENDA E EMPREGO	VENCIMENTO	APRESENTACAO	VALIDADE PRODUTO	CATEGORIA	ASSUNTO PETICAO	EXPEDIENT E PETICAO	VERSAO
RESOLUÇÃO-RE Nº 5.465, DE 30 DE DEZEMBRO DE 2020	PORTOMAGGIOR E INDÚSTRIA E COMÉRCIO DE PRODUTOS QUÍMICOS LTDA - EPP	3.03086-2	MULTIMIX GEL	25351.700694/2009-23	3.3086.0005.001-8	PRODUTO DE USO PROFISSIONAL OU DE VENDA RESTRITA	03/2030	SERINGA PLASTICA + CAIXA DE CARTOLINA	24 Meses	3206025 INSETICIDA PARA EMPRESAS ESPECIALIZADAS	334 REG. SANEANTES - Revalidação de Registro	2091244/19-7	