**Tarefa 1**

Realizar a raspagem dos dados da página [dados.gov.br](https://dados.gov.br/dados/reuso/1) e a extração da base: Boletim de Arrecadação dos Tributos Estaduais.

**Objetivo:** Processar os dados das abas indicadas, pulando a primeira linha do arquivo e carregando os dados de forma normalizada no banco de dados.

**1. Automatizar a extração da planilha**

* Construa um script onde seja realizada a extração da planilha, clicando em "ver. recursos" -> "Link".

**2. Normalizar os dados da planilha extraída.**

* Processar as abas:
  + uf
  + Arrecadação por setor
  + Arrecadação CNAE
  + Contribuinte por UF
* **Pule a primeira linha de cada aba** ao importar os dados, com exceção da aba "uf".

**3. Inserir os dados no banco de dados.**

* Utilize um banco de dados local (PostgreSQL ou SQLite).
* Cada aba deve corresponder a uma tabela separada no banco, com os nomes:
  + uf
  + arrecadacao\_setor
  + arrecadacao\_cnae
  + contribuinte\_uf

**4. Observações**

* Os dados de todas as abas devem ser processados e carregados no banco.
* Garanta que a conexão com o banco de dados seja reutilizável e documente como configurá-la.
* Adicione uma verificação para garantir que não existem duplicatas ao carregar os dados no banco.

**5. Requisitos Adicionais**

* Documentação:
  + Explique no README.md como configurar o ambiente, rodar o script e verificar os dados no banco.
* Entrega:
  + Código Python no formato .py (não notebooks).
  + Um dump do banco de dados ou instruções para criar as tabelas e validar os dados.

**Tarefa 2**

1. **Extração das Bases**
   1. Automatizar o download das planilhas disponíveis no site da ANP.
   2. Garanta que todas as abas de cada planilha sejam processadas.
2. **Preparação dos Dados**
   1. Processar as planilhas extraídas e integrá-las com a base fornecida anteriormente (arquivo dados.gov.br).
   2. Realizar o seguinte processamento:
      1. **Unir os dados das bases da ANP (2022 e 2023).**
      2. Integrar os dados da ANP com as bases extraídas anteriormente (primeira atividade).
   3. **Normalizar os dados finais:**
      1. Renomeie as colunas para padronizá-las, garantindo clareza e consistência.
      2. Exemplos de nomes finais (sugestão):
         1. ano, estado, producao\_petroleo, producao\_gas, setor, valor\_arrecadado, uf, quantidade\_contribuintes.
3. **Carga no Banco de Dados**
   1. Escolher entre SQLite ou PostgreSQL para armazenar os dados.
   2. Criar uma única tabela consolidada no banco de dados chamada dados\_integrados.

#### **Observações**

* **Automatização do Download:** O script deve ser capaz de realizar o download diretamente do site da ANP e salvar as planilhas em um diretório local.
* **Tratamento de Dados Faltantes:**
  + Valores ausentes ou inconsistentes devem ser tratados (e.g., preenchidos com NULL ou valores padrões).
* **Normalização:**
  + Ajustar os tipos de dados para manter consistência, como datas no formato YYYY-MM-DD, valores numéricos com separadores adequados, etc.
* **Documentação:**
  + Criar um README.md detalhando:
    - Como executar a automação de download.
    - Como rodar o pipeline de processamento e carga.
    - Informações adicionais, como dependências instaladas e estrutura final da tabela.