



Centro Universitário de Brasília (CEUB)

Faculdade de Tecnologia e Ciências Sociais Aplicadas (FATECS)

Thales Rassi Porto de Matos - 22400186

Gabriel Marques da Rocha - 22451254

Gabrielle Gutierrez - 22350026

Pedro Klein - 22105154

Matheus de Moraes - 22352763

Henrique Lessa - 22402204

Documentação ETL – TIPO EQUIPAMENTO

Brasília

2025

Thales Rassi Porto de Matos

Gabriel Marques da Rocha

Gabrielle Gutierres

Pedro Klein

Matheus de Moraes

Henrique Lessa

Documentação ETL – TIPO EQUIPAMENTO

Atividade final apresentada à Faculdade de
Tecnologia e Ciências Sociais Aplicadas
(FATECS)

, do Centro Universitário de Brasília
(CEUB) como parte integrante do currículo
da disciplina Interação Humano
Computador, da graduação em Ciência da
computação

Professora responsável: Kadidja Valeria
Reginaldo de Oliveira

Brasília

2025

SUMÁRIO

IDENTIFICAÇÃO	04
OBJETIVO DO ETL	04
FONTES DE DADOS DE ENTRADA	04
TABELAS DE DESTINO IMPACTADAS	05
FLUXO DO ETL	05
MAPEAMENTO DE CAMPOS (ORIGEM → DESTINO)	07
DEPENDÊNCIAS E PRÉ-REQUISITOS	07
COMO EXECUTAR	07

IDENTIFICAÇÃO

Script: ETL_TIPO_EQUIPAMENTO.py

Responsável(is): Equipe Radiologia DF

OBJETIVO DO ETL

Extrair, padronizar e consolidar os dados referentes aos tipos de equipamentos de diagnóstico por imagem existentes no Distrito Federal, calcular indicadores derivados (público, privado, funcionando, parado) e inserir ou atualizar esses dados na tabela tipo_equipamento.

O propósito é garantir que a dimensão de tipos de equipamento esteja completa, com quantidades atualizadas e consistente com os dados oficiais do SUS.

FONTES DE DADOS DE ENTRADA

Arquivo principal:

1. dirty_data_qtd equip_img_SUS_por_tipo.csv
 - Caminho esperado: diretório local do projeto
 - Formato: CSV
 - Separador: padrão (“,”)
 - Encoding: padrão do pandas
 - Periodicidade: conforme atualização dos dados do SUS referentes ao inventário de equipamentos por tipo

TABELAS DE DESTINO IMPACTADAS

1. tipo_equipamento
 - Tabela ontendo informações de cada tipo de equipamento.

FLUXO RESUMIDO DO ETL

Passo 1 – Tratamento do dataset bruto

- Para cada item em DATASETS_RA:
 - Valida se o arquivo existe; se não existir, exibe aviso e ignora.
 - Lê o CSV com os parâmetros adequados (sep, skiprows, skipfooter).
 - Valida se a coluna indicada em coluna_ra existe.
 - Extrai a coluna de RA:
 - Remove valores nulos
 - Converte para string
 - Aplica strip
 - Remove entradas vazias
 - Converte para formato Title Case
 - Concatena todas as listas de RAs extraídas.
 - Remove duplicidades.
 - Ordena alfabeticamente.
 - Retorna um DataFrame final contendo apenas uma coluna: nome_ra

Passo 2 – Carregamento no banco de dados

- Busca todos os tipos de equipamento existentes na tabela tipo_equipamento, criando um dicionário: { nome_lowercase : id_tipo_equipamento }
- Para cada linha do dataset tratado:
 - Se o nome já existir:
 - Adiciona um registro à lista de atualizações
 - Se não existir:
 - Adiciona um registro à lista de inserções
- Caso haja inserções:
 - INSERT INTO tipo_equipamento (nome, descricao, quantidade_publico, quantidade_privado, quantidade_funcionando_sus, quantidade_parado_sus) VALUES %s
- Caso haja atualizações:
 - UPDATE tipo_equipamento SET quantidade_publico = %s, quantidade_privado = %s, quantidade_funcionando_sus = %s, quantidade_parado_sus = %s WHERE id_tipo_equipamento = %s
 - Realiza commit e imprime quantos tipos de equipamento foram inseridos e quantos foram atualizados.

Passo 3 – Encerramento

- Fecha o cursor
- Fecha a conexão com o banco
- Imprime mensagem indicando conclusão do ETL

MAPEAMENTO DE CAMPOS (ORIGEM → DESTINO):

Origem no CSV	Destino na tabela tipo Equipamento
equipamento → string	nome
existentes_SUS	quantidade_publico
existentes – existentes_SUS	quantidade_privado
em_uso_SUS	quantidade_funcionando_sus
existentes_SUS – em_uso_SUS	quantidade_parado_sus

Observação: O campo descricao não é preenchido e permanece como NULL.

DEPENDÊNCIAS E PRÉ-REQUISITOS

Bibliotecas Python:

- os
- pandas
- psycopg2
- psycopg2.extras
- python-dotenv (load_dotenv)
- config_db (função get_conn para conexão com o banco)

Tabelas:

- tipo_equipamento (já existente)

Requisitos de dados

- Arquivo dirty_data_qtd_equip_img_SUS_por_tipo.csv presente no caminho esperado
- Colunas obrigatórias presentes e com tipos coerentes

COMO EXECUTAR

- Garantir que o ambiente virtual (se houver) esteja ativado.
- Garantir que o arquivo .env esteja configurado com os parâmetros de conexão ao banco.
- Garantir que o CSV está presente no caminho indicado pela constante DATASET_EQUIP_TIPO
- Executar: python3 ETL_TIPO_EQUIPAMENTO.py

Pré-condições:

- Banco PostgreSQL acessível
- Tabela tipo_equipamento criada
- Dataset limpo e compatível