**Processo de ETL para Análise de Casos de Tuberculose**

**Introdução**

Este processo de ETL (Extract, Transform, Load) foi desenvolvido em Python com objetivo de extrair, transformar e carregar dados relacionados aos casos de tuberculose no estado de Goiás. O objetivo é preparar os dados para análise, garantindo que estejam limpos, filtrados e agregados de forma adequada.

**Etapas do Processo**

1. **Extração dos Dados**
   * **Fontes de Dados**: Os dados são extraídos de arquivos CSV localizados em diretórios específicos.
     + origem.csv: Contém os dados brutos dos casos de tuberculose.
     + d\_municipio.csv: Contém informações sobre os municípios.
     + d\_tempo.csv: Contém informações sobre o tempo.
2. **Transformação dos Dados**
   * **Filtragem Inicial**: Os dados são filtrados com base em critérios específicos:
     + tp\_entrada deve ser 1, 4 ou 6.
   * 1 = Caso Novo
   * 4 = Não Sabe
   * 6 = Pós-óbito
     + tp\_forma deve ser 1 ou 3.
   * 1 = Pulmonar
   * 3 = Pulmonar + Extrapulmonar
     + tp\_situacao\_encerramento não deve ser 6.
   * 6 = Mudança de Diagnóstico
     + dt\_diagnostico\_sintoma deve ser a partir de 1º de janeiro de 2021.
   * **Junção com Dados de Municípios**: Os dados filtrados são combinados com a tabela de municípios para incluir o nome completo dos municípios de Goiás.
   * **Conversão de Datas**: A data de diagnóstico é convertida para o formato “Mês de Ano”.
   * **Agregação dos Dados**: Os dados são agrupados por município e mês, e a quantidade de casos é contada.
3. **Carga dos Dados**
   * **Criação de um DataFrame Completo**: Um DataFrame é criado para garantir que todos os municípios e meses estejam presentes, mesmo que não haja casos registrados.
   * **Combinação Final**: O DataFrame completo é combinado com os dados agregados, preenchendo valores ausentes com zero.
   * **Salvamento dos Dados**: O resultado final é salvo em um arquivo CSV (saida.csv).