

Estrutura do Exercício de ETL com Python e Power BI

Este exercício tem como objetivo proporcionar uma experiência prática no desenvolvimento de um **pipeline de ETL (Extração, Transformação e Carga)**, integrando **Python, banco de dados relacional e Power BI**. O trabalho será dividido em **quatro etapas principais**:

◊ 1. Extração (E)

Obter dados a partir de **fontes públicas** em diferentes formatos, como **CSV, APIs ou bancos de dados abertos**.

◊ 2. Transformação (T)

Utilizar **Python** (bibliotecas como *pandas* e *numpy*) para:

- Limpeza de dados (remoção de nulos e duplicados).
- Padronização de colunas (strings, datas, categorias).
- Criação de novas variáveis derivadas.
- Categorização de variáveis e agrupamentos.
- Geração de estatísticas descritivas (média, mediana, desvio padrão).

◊ 3. Carga (L)

Armazenar os dados tratados em um **banco de dados local** (SQLite ou PostgreSQL), garantindo:

- Estrutura organizada em tabelas.
- Definição de chaves primárias e estrangeiras para modelagem relacional.

◊ 4. Análise e Visualização

Conectar o **Power BI** ao banco de dados para criar **dashboards interativos e dinâmicos**, respondendo a perguntas de negócio e permitindo exploração dos dados por meio de filtros e gráficos visuais.

Fontes de Dados Sugeridas

Os datasets podem ser obtidos em repositórios abertos e confiáveis, como:

- [Kaggle](#)
- [UCI Machine Learning Repository](#)
- [Data.gov](#)
- [IBGE](#)

Exemplos práticos de datasets interessantes:

- Desempenho e produtividade de empregados → [Kaggle Dataset](#)
- Retenção de profissionais (RH) → [Kaggle Dataset](#)
- Trabalho remoto e saúde mental → [Kaggle Dataset](#)
- Queimadas nas florestas brasileiras → [Kaggle Dataset](#)
- Crimes no Brasil → [Kaggle Dataset](#)
- Degradação da floresta amazônica → [Kaggle Dataset](#)
- Séries temporais de temperaturas em cidades brasileiras → [Kaggle Dataset](#)

Conceitos Práticos a Serem Desenvolvidos

- **Python** → ETL, limpeza e transformação de dados.
- **Banco de Dados** → Estrutura relacional, chaves e integridade.
- **Análise de Dados** → Criação de métricas e indicadores relevantes.
- **Power BI** → Dashboards dinâmicos com filtros, segmentações e visualizações interativas.