**Documentação Detalhada do Projeto de Processamento de Dados no Google Drive**

**Visão Geral**

Este projeto tem como objetivo extrair, transformar e carregar (ETL) dados armazenados em planilhas do Google Drive. Utiliza-se a API do Google Drive para acessar os arquivos e a biblioteca Pandas para a manipulação dos dados.

**1. Autenticação no Google Drive**

A classe GoogleDriveManager gerencia a autenticação na API do Google Drive e fornece métodos para interagir com arquivos.

**1.1 Autenticação**

A autenticação é realizada via uma conta de serviço utilizando credenciais armazenadas no arquivo JSON localizado em:

CREDENTIALS\_PATH = 'credentials/people-analytics-pipoca-agil-google-drive.json'

**1.2 Métodos Principais**

* get\_file\_by\_name(folder\_id\_origem, file\_name): Busca um arquivo pelo nome dentro de uma pasta no Google Drive.
* download\_excel(file\_id): Faz o download de um arquivo Excel para um fluxo de bytes.
* upload\_file(folder\_id\_origem, file\_name, file\_path): Faz o upload de um arquivo para o Google Drive.

**2. Extração de Dados**

A classe DataExtractor é responsável por baixar os arquivos Excel do Google Drive.

**2.1 Método Principal**

* load\_excelfile(): Busca o arquivo no Google Drive e carrega os dados em um DataFrame Pandas.

**3. Transformação de Dados**

A classe DataTransformer é responsável por limpar e padronizar os dados.

**3.1 Principais Funcionalidades**

* **Normalização de Texto:**
  + retirar\_acento(frase): Remove acentos de uma string.
  + normalizar\_perguntas(pergunta): Converte perguntas para um formato padronizado.
* **Validação e Padronização:**
  + verificar\_email(email): Valida e-mails.
  + padronizar\_datastring(dataStamp): Converte datas para o formato dd-mm-yyyy HH:MM:SS.
  + renomear\_colunas\_autoavaliacao(): Renomeia colunas da planilha "autoavaliação".
  + validar\_email(): Remove e-mails inválidos.
  + adicionar\_caderno\_pergunta(): Adiciona uma coluna identificando o tipo de questionário.
  + clean\_empty\_rows(): Remove linhas duplicadas.
* **Transformação Principal:**
  + transformar\_dados(): Aplica todas as transformações ao DataFrame.

**4. Processo Principal (ETL)**

A função processar\_arquivo executa todas as etapas do processo ETL.

**4.1 Fluxo de Processamento**

1. **Extração:** Baixa o arquivo do Google Drive.
2. **Transformação:** Aplica limpeza, normalização e validação de dados.
3. **Carga:** Salva o arquivo transformado no Google Drive.

**4.2 Código Principal**

folder\_id\_origem = "1E6AEUGqRp3IJsWV4qAwMRJK\_tMD7wDYT" # Pasta de origem (RAW)

folder\_id\_destino = "1WJlq1C\_uLq9J3Ta-lVAkQVv7AzblftsD" # Pasta de destino (TRUSTED)

file\_name = "autoavaliacao.xlsx"

processar\_arquivo(folder\_id\_origem, folder\_id\_destino, file\_name)

**5. Estrutura de Pastas no Google Drive**

* people\_analytics/raw/: Contém os arquivos brutos extraídos de questionários.
* people\_analytics/trusted/: Contém os arquivos transformados e validados.

**6. Considerações Finais**

Este projeto automatiza o processamento de dados de autoavaliação, garantindo qualidade e padronização das informações armazenadas. Caso novos questionários sejam adicionados, o código pode ser facilmente ajustado para suportá-los.