

```
import pandas as pd
import pyarrow as pa
import pyarrow.parquet as pq

def extrair(origem: str) -> pd.DataFrame:
    df = pd.read_csv(origem, encoding='utf-8', sep=',')
    return df

def tratar(df: pd.DataFrame) -> pd.DataFrame:
    # 1. Ajuste de valores nulos
    df = df.fillna({
        'col_numérica': 0,
        'col_texto': 'desconhecido'
    })

    # 2. Renomeação de variáveis
    df = df.rename(columns={
        'colAntiga1': 'colNova1',
        'colAntiga2': 'colNova2'
    })

    # 3. Ajuste de tipos
    df['colNova1'] = df['colNova1'].astype(int)
    df['data'] = pd.to_datetime(df['data'], format='%Y-%m-%d')

    # 4. Padronização de códigos institucionais
    df['cod_instituicao'] = df['cod_instituicao'].str.zfill(6)

    return df

def salvar_parquet(df: pd.DataFrame, path: str):
    table = pa.Table.from_pandas(df, preserve_index=False)
    pq.write_table(table, path, compression='snappy')

def main():
    df = extrair('dados/brutos/input.csv')
    df = tratar(df)
    salvar_parquet(df, 'dados/tratados/saida.snappy.parquet')

if __name__ == '__main__':
    main()
```