

## Python

```
from airflow.decorators import dag, task
from datetime import datetime
import pandas as pd

@dag(
    dag_id="etl_pipeline",
    start_date=datetime(2023, 1, 1),
    schedule_interval=None,
    catchup=False,
    tags=["example", "etl"]
)
def etl_pipeline_dag():
    @task
    def extract():
        # NOTE: This path is specific to your local machine and may need to be changed
        # to a path accessible by your Airflow worker for a real production DAG.
        df=pd.read_csv('C:/Users/nikhi/airflow/airflow-docker/dags/Handle_Missing_Data.csv')
        print(f'[Extract] Raw shape: {df.shape}')
        return df.to_dict()

    @task
    def transform(data):
        df = pd.DataFrame(data)
        # Simple transformation: dropping all rows with any missing values
        df_clean = df.dropna()
        print(f'[Transform] Cleaned shape: {df_clean.shape}')
        return df_clean.to_dict()

    @task
    def load(data):
        df = pd.DataFrame(data)
        print(f'[Load] Final Cleaned Data: \n')
        print(df.head())

    # Define the task dependency flow
    load(transform(extract()))
```

```
etl_pipeline_dag()
```