"""

DAG: upload_to_google_sheets_dag

Opis: Ten DAG obsługuje proces wczytywania danych o obrazach i ich metadanych, dzielenie danych na zbiory treningowe i testowe,

oraz zapisywanie wynikowych zbiorów do arkuszy Google Sheets.

Kroki pracy:

1. **load_data_task**:

- Ładuje dane o obrazach i ich adnotacjach z lokalnych folderów `/images` i `/annotations`.

- Tworzy DataFrame zawierający trzy kolumny: `breed` (rasa psa), `image_path` (ścieżka do obrazu),

oraz `annotation_path` (ścieżka do adnotacji).

- Zapisuje dane w formacie słownika w XCom, aby mogły być użyte przez kolejne zadania.

2. **split_data_task**:

- Pobiera dane z `load_data_task` za pomocą XCom.

- Konwertuje dane z formatu słownika na DataFrame.

- Dzieli dane na dwa zbiory: treningowy (70%) i testowy (30%), korzystając z funkcji `train_test_split`.

- Zapisuje oba zbiory jako słowniki w XCom.

3. **upload_train_data_to_sheets**:

- Pobiera dane treningowe z XCom.

- Formatuje je jako DataFrame.

- Zapisuje dane do arkusza Google Sheets w zakładce `Train Data`.

4. **upload_test_data_to_sheets**:

- Pobiera dane testowe z XCom.

- Formatuje je jako DataFrame.

- Zapisuje dane do arkusza Google Sheets w zakładce `Test Data`.

Struktura folderów:

- `/opt/airflow/dags/images/Images` – folder zawierający obrazy poszczególnych ras psów.

- `/opt/airflow/dags/annotations/Annotation` – folder zawierający metadane obrazów w formacie XML.

Wymagania:

- Klucz API Google Sheets (z uprawnieniami do zapisu w arkuszu).

- Zainstalowane i skonfigurowane zależności, takie jak `pandas`, `requests`, i `scikit-learn`.

Założenia:

- Arkusz Google Sheets istnieje i ma odpowiedni `sheet_id`.

- Nazwy arkuszy w arkuszu Google (`Train Data` i `Test Data`) są zdefiniowane i dostępne.

Zależności między zadaniami:

- `load_data_task` → `split_data_task` → `[upload_train_data_to_sheets, upload_test_data_to_sheets]`