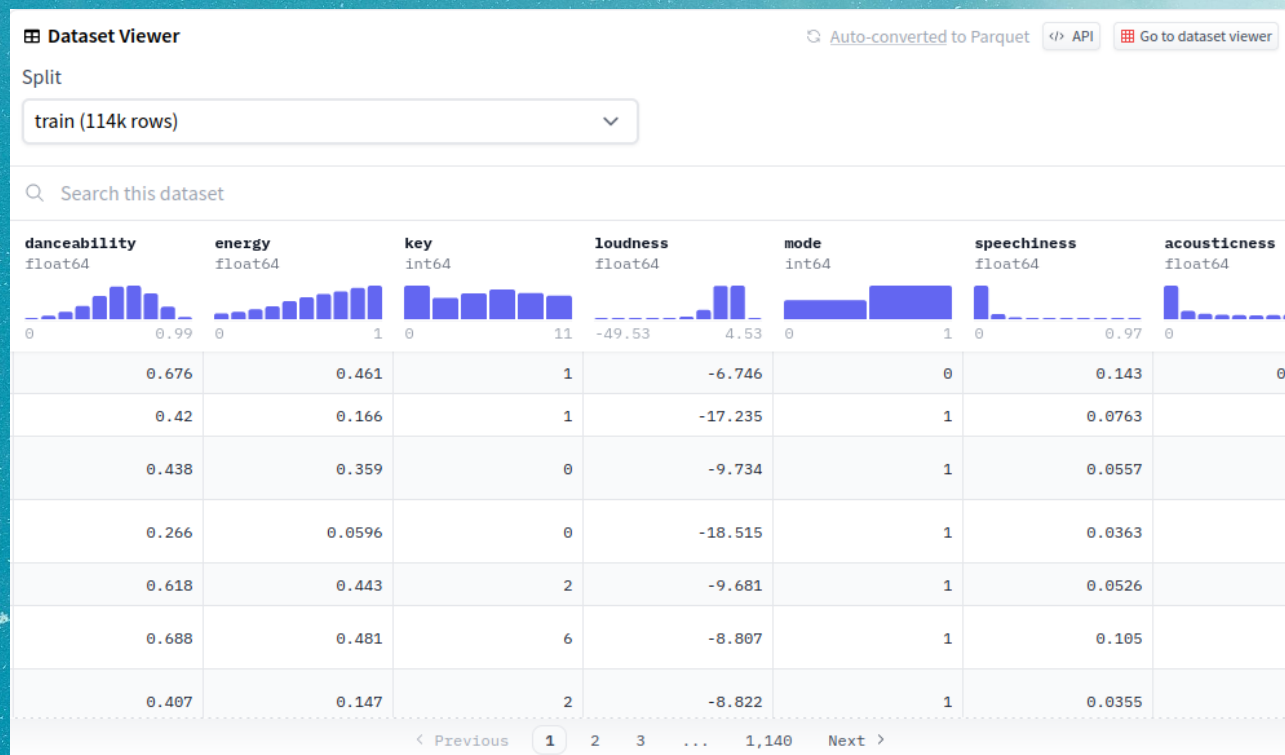


PROJEKT ADZD

APACHE AIRFLOW

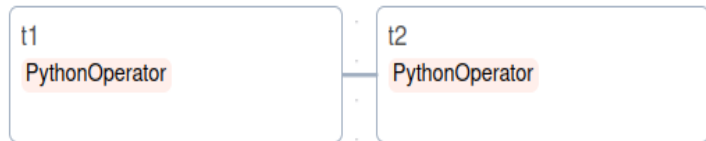
Dataset

- <https://huggingface.co/datasets/maharshipandya/spotify-tracks-dataset>
- 114,000 piosenek z opisanymi cechami



Wersja podstawowa

DAG Summary		
Total Tasks		2
PythonOperators		2



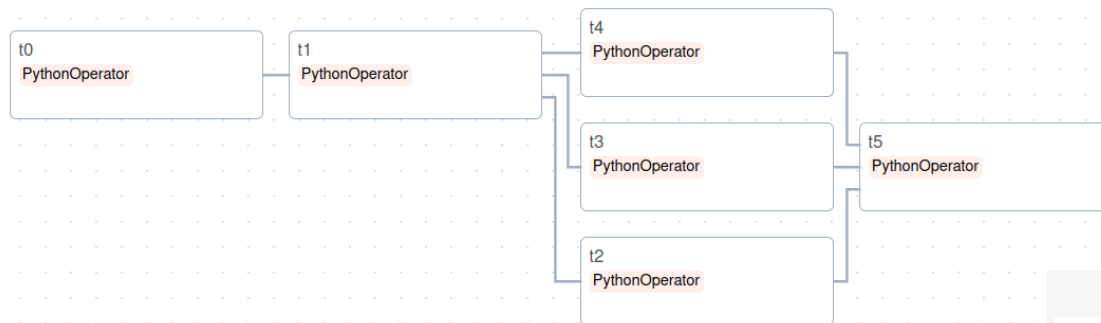
Ten DAG ma dwa zadania –

- T1 – wczytuje dane
- T2 – wykonuje obliczenia: dzieli dane na trzy zbiory według obliczonych progów cech „loudness”, w każdym z tych zbiorów grupuje dane po gatunku muzycznym i zlicza ile jest utworów danym zakresie loudness o określonym gatunku. Na koniec zestawia razem te wyniki.

Wersja ze równoległymi obliczeniami

DAG Summary

Total Tasks	6
PythonOperators	6



Ten DAG jest rozbudowaną wersją poprzedniego – zadanie T2 zostało rozbite na pięć podzadań, z których 3 można wykonywać równocześnie.

Wyniki eksperymentów

Wersja podstawowa

Duration

00:00:15

Wersja rozszerzona

Duration

00:00:43

Eksperyment podstawowy przeprowadzono szybciej niż ten, który obejmował równoległe zadania. Może to wynikać z faktu, że zużycie czasu wynika głównie z dostępu do bazy danych, a nie z obliczeń. Widzimy kilka sposobów rozwiązania tego:

- zdobyć lub stworzyć znacznie większy zestaw danych
- dodać znacznie więcej obliczeń do tasków
- znaleźć szybszy sposób zapisywania i wczytywania danych

Dziękujemy

Agnieszka Szynalik, Elżbieta Wierciak