

# PROJEKT ADZD

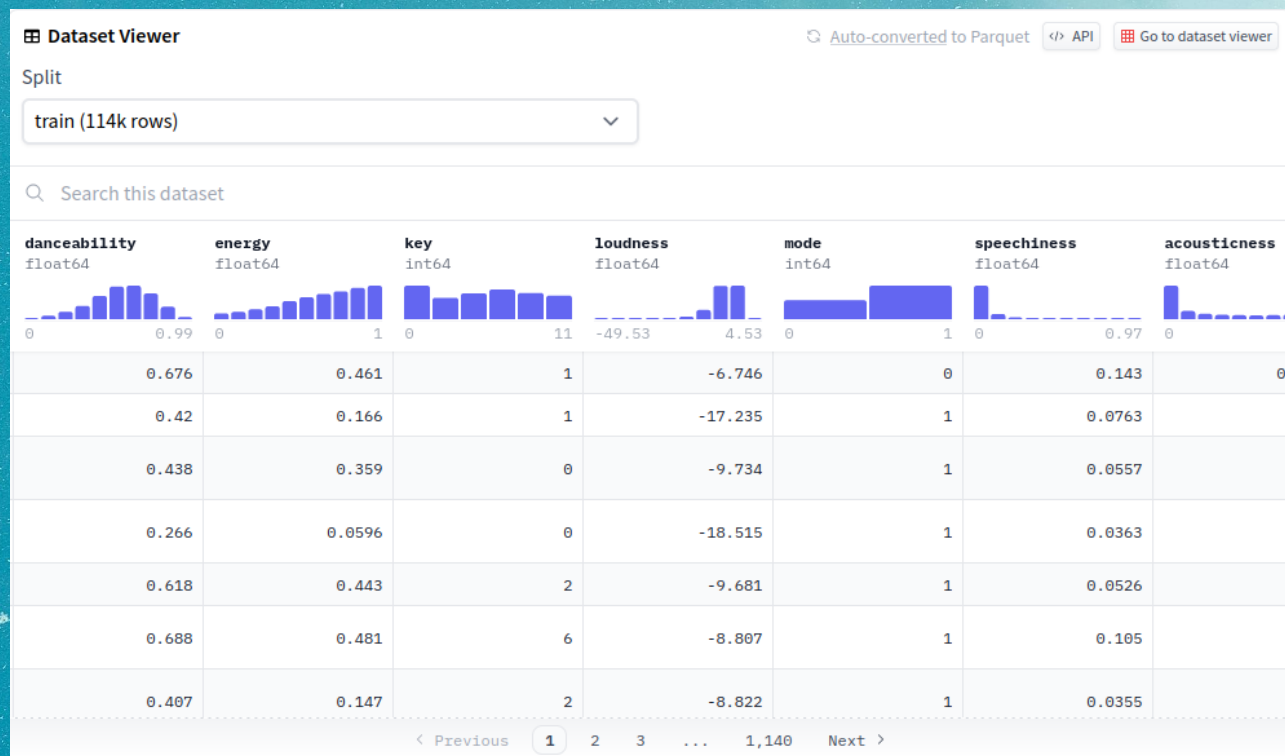
---

APACHE AIRFLOW



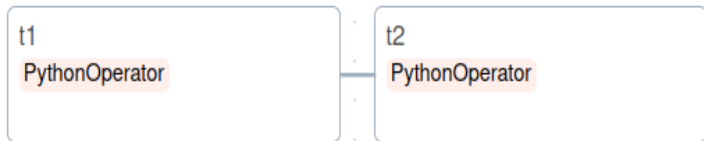
# Dataset

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



# Wersja podstawowa

DAG Summary	
Total Tasks	2
PythonOperators	2



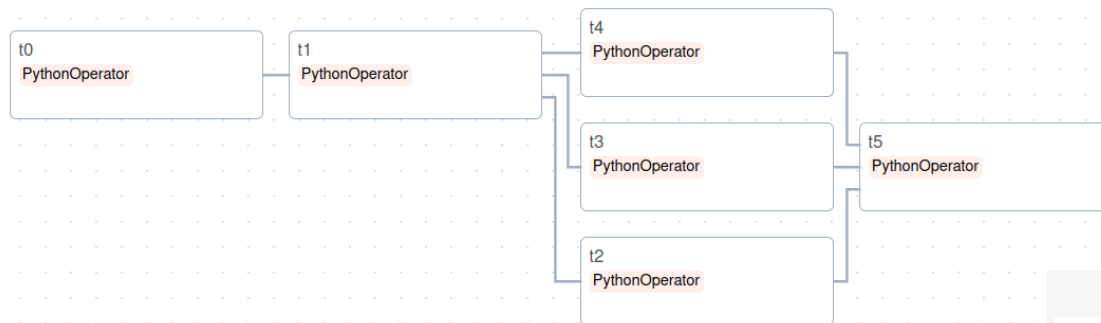
Ten DAG ma dwa zadania –  
T1 wykonuje wszystkie obliczenia,  
podczas gdy T2 jest funkcją końcową



# Wersja ze równoległymi obliczeniami

## DAG Summary

Total Tasks	6
PythonOperators	6



T1 z poprzedniego DAG został podzielony na kilka zadań (które są podzbiorem T1)



# Wyniki eksperymentów

Duration

00:00:15

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