

# **Tritone Assignment - Detailed Summary**

Submitted by: Deepak Kumar

Artist Chosen: Coldplay

## **Personal Details:**

Name: Deepak Kumar

Mobile: 7479496478

Email: deepakkumar2.mcac2024@ritroorkee.com

## **Project Summary:**

This project involved working with a large unclaimed musical works dataset (~7GB TSV) and a chosen artist's catalog from Spotify. The main objective was to cross-reference tracks using ISRC codes, identify unclaimed works, and document the approach, tools used, and observations in a structured and professional manner.

Key objectives:

- Efficiently process large datasets without overloading memory.
- Standardize and clean ISRC codes for accurate matching.
- Maintain a professional summary for submission.

## **Approach & Steps:**

- Connected to Spotify API (Spotipy) to fetch the full catalog of the chosen artist including track name, album, release date, and ISRC codes.
- Processed the unclaimedmusicalworkrightshares.tsv dataset in chunks to avoid memory issues.
- Standardized all ISRC codes to ensure accurate cross-referencing.
- Documented the process and decisions taken at each step.
- Prepared a final submission-ready summary without including the actual track or match data.

## **Tools & Libraries Used:**

- Python 3.10+
- Pandas (for TSV processing and data handling)
- Spotipy (Spotify API access)
- OpenPyXL (Excel output)
- DuckDB (efficient large dataset handling)
- Microsoft Excel / CSV (data review and verification)

## **Observations / Notes:**

- Large dataset processing was successfully handled using chunked reading.
- ISRC codes were cleaned and standardized to maintain consistency.
- The chosen tools enabled efficient data retrieval, processing, and documentation.
- All steps and decisions have been documented clearly for professional review.
- Summary is concise, submission-ready, and follows Tritone's instructions.

**Thank You!**