

# Max Branca

2061432

I propose implementing a comprehensive data warehouse utilizing four complementary music datasets:

- Song Features Dataset (<https://www.kaggle.com/datasets/ayushnitb/song-features-dataset-regressing-popularity>), derived from the official Spotify Web API, which provides extensive song metadata including audio attributes (danceability, energy, acousticness), artist information, and popularity metrics spanning multiple years of music releases;
- Spotify 1.2M+ Songs Dataset (<https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs>), which offers extensive genre classifications for millions of tracks, contains additional popularity metrics and provides expanded temporal coverage across multiple decades;
- Billboard Hot 100 Songs Dataset (<https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs>), which presents the ranking of the Top 100 most popular songs in the United States published weekly since 1958 and incorporates commercial performance metrics based on sales, radio play, and online streaming.
- Worldwide Music Artists Dataset (with image) (<https://www.kaggle.com/datasets/harshdprajapati/worldwide-music-artists-dataset-with-image>), which contains general information about artists, such as name, genre and country.

This project aims to create a comprehensive music analytics data warehouse that will enable multidimensional analysis of music trends, audio characteristics, and commercial performance across different time periods. By integrating these four datasets, I will develop a robust dimensional model capable of supporting complex analytical queries (e.g. how audio attributes of popular music have changed across different decades, how successful artists adapt their musical style over their careers, which audio characteristics most strongly correlate with commercial success).