**Project Proposal**

Group members:

- Jacopo Maria Barile 2208592

Chosen type of project:

[**DW**] Select a set of data sources, integrate them by means of ETL operations (you can either use an integration tool or write simple scripts to solve common ETL problems), define the DFM model for your data, define and populate the corresponding star schema or snowflake schema, by using a relational DB (e.g., Postgres) .

**Focus**: ability to combine data from different datasets in order to produce a relevant unified source of analysis.

Description of the dataset which you intend to use:

- IMDB Non-Commercial Datasets (IMDB datasets link)
  In the previous link there are 3 datasets of interest:
  - Title.basics.tsv.gz that includes all the title info such as type, genre, duration and start year.
  - Title.ratings.tsv.gz that includes the average rating and total number of votes per each title.
  - Name.basics.tsv.gz that includes actors and artists basic info, plus the main titles they are known to.
- Top 1000 most played Spotify songs of all time (Kaggle top 1000 Spotify songs)
  This dataset contains track name, artist, album and release date of the songs.

Brief description of the work you intend to do:

The idea is to integrate multiple entertainment-related datasets using an ETL (Extract, Transform, Load) process in order to analyze correlations between highly rated movies and popular songs.

The integration process will involve cleaning and transforming the data to create connections between songs and movies (e.g., identifying songs created for movies), and mapping actors to the movies they appear in. The goal is to construct a unified data warehouse, organized in a star schema, that supports analytical queries to explore:

- Whether there is a relationship between popular songs and the ratings of the movies they are associated with.
- What's the average rating of titles that include famous artists.
- Which actors most frequently appear in highly rated movies.
- General patterns connecting music popularity and movie success.