



Spotify®

100 Years of
Music

Data Contents

- **tracks.csv** (audio features of tracks, 600k rows, 20 attributes)
- **artists.csv** (popularity metrics of artists, 1.1M rows, 5 attributes)

artist.csv

- - id (Id of artist)
- - name (Name of artist)
- - followers (Total number of followers of artist)
- - popularity (Popularity of given artist based on all his/her tracks)
- - genres (Genres associated with this artist)

tracks.csv

Numerical Features

- - acousticness (Ranges from 0 to 1)
- - danceability (Ranges from 0 to 1)
- - energy (Ranges from 0 to 1)
- - duration_ms (Integer typically ranging from 200k to 300k)
- - instrumentalness (Ranges from 0 to 1)
- - valence (Ranges from 0 to 1)
- - popularity (Ranges from 0 to 100)
- - tempo (Float typically ranging from 50 to 150)
- - liveness (Ranges from 0 to 1)
- - loudness (Float typically ranging from -60 to 0)
- - speechiness (Ranges from 0 to 1)

Dummy

- - mode (0 = Minor, 1 = Major)
- - explicit (0 = No explicit content, 1 = Explicit content)

Categorical

- - key (All keys on octave encoded as values ranging from 0 to 11, starting on C as 0, C# as 1 and so on...)
- - timesignature (The predicted timesignature, most typically 4)
- - artists (List of artists mentioned)
- - artists (Ids of mentioned artists)
- - release_date (Date of release mostly in yyyy-mm-dd format, however precision of date may vary)
- - name (Name of the song)

Goals:

Predict popularity rating

On the basis of current data, analysis of which features play the most important role in determining the popularity of the song. Divide the popularity into 10 categories. And determine the rating outcome based on the other features.

Trend analysis

Perform time-series analysis on "track.csv" data. Work on visualizations or statistical analysis and even a forecast program for upcoming tracks in the next ten years.

Differentiate genres

The fact that there are too many genres to fully understand, so a more generalized approach was needed: Creating super-genres out of genres! In this task, you are expected to understand how clustering works, how a prediction is made, and the evaluation methods.

The background features three large, semi-transparent circles with a gradient from dark purple to bright blue. One circle is positioned at the top left, another at the bottom left, and a third, larger one is centered on the right side.

Thanks !