

Project 1

This dataset comprises Spotify tracks spanning across 125 diverse genres. Every track includes specific audio characteristics. Presented in CSV format, the data is organized in a table-like structure, ensuring swift loading.

Link to dataset:

<https://drive.google.com/drive/folders/1r99Xk68Hc-bRFNx2Hr1cAEDKUcGTP5M?usp=sharing>

Usage:

Potential applications of this dataset include:

1. Creating a Recommendation System tailored to user preferences or inputs.
2. Classifying tracks using their audio features and the range of genres they cover.
3. Any other innovative use you can conceive. Suggestions and discussions are welcome!

Column Description:

1. track_id: Unique Spotify identifier for each track.
2. artists: Names of artists involved in the track, separated by ';' for multiple artists.
3. album_name: Title of the album containing the track.
4. track_name: Title of the individual track.
5. popularity: A score from 0 to 100 indicating the track's popularity, where 100 is the most popular. This score is algorithmically determined, primarily based on the track's play count and the recency of these plays. A track's current play frequency influences its popularity more than past plays. Tracks appearing in multiple forms (like in an album and a single) have separate ratings. Note that artist and album popularity are also derived from track popularity.
6. duration_ms: Length of the track in milliseconds.
7. explicit: Indicates if the track contains explicit lyrics ('true' for explicit content; 'false' for no explicit content or if it's unknown).
8. danceability: A metric ranging from 0.0 (least danceable) to 1.0 (most danceable), assessing a track's suitability for dancing based on tempo, rhythm, beat strength, and general regularity.
9. energy: A perceptual measure ranging from 0.0 to 1.0, gauging the track's intensity and activity. Tracks that are fast, loud, and noisy are considered high energy, like death metal, whereas a Bach prelude would be low energy.
10. key: The musical key of the track, represented by integers following standard Pitch Class notation (e.g., 0 = C, 1 = C#/D ♭, 2 = D). A value of -1 indicates an undetected key.

11. loudness: Measures the average loudness of the track in decibels (dB).
12. mode: Indicates the track's modality, with 1 for major mode and 0 for minor mode, determining the type of scale that forms its melodic basis.
13. speechiness: Assesses the extent of spoken words in a track. Values near 1.0 suggest a predominance of speech (like talk shows or audio books). Scores above 0.66 typically indicate tracks composed entirely of spoken words. Those between 0.33 and 0.66 may include a mix of music and speech, such as in rap music. Scores below 0.33 generally represent music or non-speech tracks.
14. acousticness: A scale from 0.0 to 1.0 indicating the likelihood of the track being acoustic, with 1.0 signifying high confidence in its acoustic nature.
15. instrumentalness: Estimates the absence of vocals in a track. Vocal-like sounds ("ooh" and "aah") are considered instrumental, whereas rap or spoken words are categorized as vocal. Values closer to 1.0 suggest a higher probability of the track lacking vocal content.
16. liveness: Detects the presence of a live audience in the recording. Higher values suggest a greater chance that the track was performed live, with values above 0.8 strongly indicating a live performance.
17. valence: A metric ranging from 0.0 to 1.0, describing the track's emotional tone. High valence tracks sound more positive (happy, cheerful, euphoric), while low valence tracks convey more negative emotions (sad, depressed, angry).
18. tempo: The track's overall estimated tempo, measured in beats per minute (BPM). Tempo in music refers to the speed or pace of a piece, derived from the average duration of a beat.
19. time_signature: Provides an estimated time signature for the track, which is a musical notation indicating the number of beats in each bar (or measure). This value varies between 3 to 7, representing time signatures from 3/4 to 7/4.
20. track_genre: Specifies the genre classification of the track.

Sources and Methodology

The data was collected and cleaned using Spotify's Web API and Python.

Project 2

Link to assignment

<https://datacatalog.worldbank.org/search/dataset/0060166>

This dataset focuses on food price inflation, a key indicator for shaping economic policy. Traditional sources of consumer prices, which often face delays and only offer aggregated data during crises, may not accurately capture price trends in rural or impoverished areas. These areas are home to large populations living in fragile conditions.

The dataset comprises estimates of food prices, aiming to provide a deeper understanding of price trends that go beyond the scope of conventional measurement methods. These estimates are derived using a machine-learning algorithm that predicts ongoing subnational price surveys, frequently achieving accuracy comparable to direct price measurements. The dataset opens up new avenues for exploring local price fluctuations in regions where people are vulnerable to localized price surges and where conventional data is unavailable.

Project 3

Build your kaggle profile

Link to competition

<https://www.kaggle.com/competitions/salary-prediction-for-job-postings>