

Data and Task Abstraction

1. Find a Dataset

[Spotify and Youtube]

(<https://www.kaggle.com/datasets/salvatorerastelli/spotify-and-youtube/data>)

전 세계 다양한 아티스트의 노래와 각 노래와 관련된 데이터셋입니다. 여기에는 노래의 음악적 특성과 두 가지 주요 플랫폼인 Spotify와 Youtube에서의 대한 변수가 포함되어 있습니다.

2. Abstract the Dataset

데이터셋에는 Spotify에서 수집한 각 노래에 대한 26개의 변수가 포함됩니다. 데이터의 수는 20716개이며, 파일의 크기는 29.3MB입니다. 데이터셋은 노래, 아티스트, Spotify 및 Youtube 지표와 관련된 정보를 가진 표 형식의 데이터셋으로 각 행은 특정 노래를 나타내고 각 열은 해당 노래와 관련된 다른 속성을 나타냅니다. 각 데이터는 다음과 같이 이루어져 있습니다.

- Nominal attribute

- Track: 노래 이름
- Artist: 아티스트 이름
- Album: 앨범 이름
- Album_type: 앨범 타입(싱글 or 앨범)
- Uri: Spotify 노래의 URI
- Url_spotify: Spotify 아티스트의 URL
- Url_youtube: YouTube 비디오의 URL.
- Title: YouTube 비디오 제목
- Channel: 비디오를 게시한 YouTube 채널 이름
- Description: YouTube 비디오 설명

- Quantitative attribute

- Key: 노래의 키(Diverging)
- Danceability: 춤추기 지표(Diverging)
- Energy: 음악 강도 지표(Diverging)
- Loudness: 전체 음악 소리 크기(dB)(Diverging)
- Acousticness: 어쿠스틱 지표(Diverging)
- Speechiness: 노래의 말로 된 부분(Diverging)
- Instrumentalness: 보컬의 존재 여부(Diverging)
- Liveness: 라이브 공연 확률(Diverging)
- Valence: 음악적 긍정성(Diverging)
- Tempo: BPM(분당 비트 수)로 나타낸 템포(Sequential)
- Duration_ms: 노래 지속 시간(밀리초)(Sequential)
- Stream: Spotify 스트림 횟수(Sequential)
- Views: YouTube 조회수(Sequential)
- Likes: YouTube 좋아요 수(Sequential)
- Comments: YouTube 댓글 수(Sequential)

키: 노래 이름인 'Track'은 곡을 고유하게 식별하는 데 사용될 수 있습니다. 하지만 같은 노래 이름이 존재하게 될 수 있어서 좋은 선택은 아닙니다.
값: 다른 모든 속성은 노래 및 비디오와 관련된 정보와 지표를 제공합니다.

3. Abstract Possible Tasks on the Dataset

3.1. 인기 아티스트 식별

아티스트별로 **Spotify** 스트리밍과 **Youtube** 조회수의 합산을 계산하여 가장 인기 있는 아티스트를 식별합니다. 이는 전체 데이터셋에서 어떤 아티스트가 가장 큰 관심을 얻고 있는지 파악하는데 도움이 됩니다.

3.2. 인기 장르 분석

장르별로 **Spotify** 스트리밍과 **Youtube** 조회수의 평균을 계산하여 각 음악 장르의 인기를 분석합니다. 이는 어떤 음악 장르가 가장 큰 관심을 받고 있는지를 이해하고 아티스트 또는 레이블이 어떤 장르로 노래를 내야 할지에 대한 결정에 도움을 줄 수 있습니다.

3.3. Youtube 동영상과 노래의 상관관계 탐색

Youtube 지표(조회수, 좋아요, 댓글 등)와 다양한 음악적 속성들 간의 상관관계를 조사합니다. 상관관계를 파악하면 어떤 음악적 특성이 두 플랫폼에서 더 높은 수치를 기록할 가능성이 높은지 알 수 있습니다. 이를 통해서 관련 플랫폼에 맞게 음악을 최적화하는 데 도움이 될 수 있습니다.

3.4. 가장 성공적인 아티스트 식별

Spotify에서 스트리밍 수와 **Youtube**에서 높은 조회수와 좋아요 수로 모두 성공을 거둔 아티스트를 식별합니다. **Spotify**, **Youtube**의 지표를 분석하여 두 플랫폼에서 모두 두드러진 존재감을 나타낸 아티스트를 식별하고 기여한 요인을 파악할 수 있습니다.

3.5. 장르 트렌드 분석

데이터셋을 탐색하여 음악적 속성들을 기반으로 주요 음악 장르를 식별합니다. 주요 속성의 분포를 분석함으로써 현재 음악 장르 트렌드를 이해할 수 있습니다. 이는 **Spotify**와 **Youtube**에서 어떤 음악 장르가 인기 있는지 파악하는 데 도움을 주며 음악 제작 및 홍보 전략을 결정하는 데 유용합니다. 또한 타임스탬프 속성이 있다고 가정했을 때 추세를 분석하여 트렌드를 발견할 수 있습니다.

3.6 아티스트와 음악 특성 간의 상관관계 분석

데이터셋을 사용하여 아티스트별로 음악 특성과 스트림 수의 상관관계를 분석합니다. 아티스트의 음악 스타일과 특성이 **Spotify**에서의 스트림 수에 어떠한 영향을 미치는지 이해할 수 있습니다. 이는 특정 음악 특성과 아티스트의 스트림 수 간의 상관관계를 보여주며 어떤 음악 특성이 중요한지 파악하는 데 도움을 줄 수 있습니다.