数据集介绍

四个数据集为5854位艺术家在100年内（1921到2020年）总共创作了98340首歌曲（包括两位及以上合作完成的情况）。

Influence\_Data介绍了这5854 名艺术家在过去90 年中的影响者和追随者。

full\_music\_data介绍了这98340首歌曲的名称 创作者姓名 创作者Id以及16个音乐方面的指标。

Data\_by\_year 介绍了这100年每年创作的音乐的特性

data\_by\_artist 介绍了这5854位每位艺术家创作的音乐的特性。

（1）“Influence\_Data”（这些数据是从网站AllMusic.com 收集来的）代表艺人自己报道的音乐影响者和追随者，以及行业专家的意见。这些数据包含5854 名艺术家在过去90 年中的影响者和追随者。

（2）“full\_music\_data”(这些数据来自Spotify 的API)提供了16 个变量条目，包括舞蹈性、节奏、响度和音调等音乐特征，以及针对98,340 首歌曲的artist\_name 和artist\_id。这些数据用于创建两个汇总数据集，包括:

a. 艺术家的平均值“data\_by\_artist”，

b. 表示跨年“data\_by\_year”。

influence\_data.csv 42770条数据

“Influence\_Data”（这些数据是从网站AllMusic.com 收集来的）代表艺人自己报道的音乐影响者和追随者，以及行业专家的意见。这些数据包含5854 名艺术家在过去90年中（1930-2010）的影响力和追随者。

前四个指标为影响者

后四个指标为追随者。

|  |  |
| --- | --- |
| influencer\_id | 给被列为影响者的人的唯一标识号。(数字或者字符串) |
| influencer\_name | 追随者或行业专家给出的有影响艺术家的姓名。(字符串) |
| influencer\_main\_genre | 最能描述有影响艺术家创作的大部分音乐的流派。(如果有的话)(字符串) |
| influencer\_active\_start | 这位有影响力的艺术家开始音乐生涯的前十年。(int 形) |
| follower\_id | 给被列为追随者的艺术家的唯一标识号。(数字或者字符串) |
| follower\_name | 追随有影响力的艺术家的艺术家的名字。(字符串)追随者的名字。 |
| follower\_main\_genre | 最能描述以下艺术家创作的大部分音乐的流派。(如果有的话)(字符串)追随者的音乐流派 |
| follower\_active\_start | 追随的艺术家音乐生涯的前十年。(整数) |

full\_music\_data.csv 98340条数据

98340首歌曲的名称 创作者姓名 创作者Id以及16个音乐方面的指标。

“full\_music\_data”(这些数据来自Spotify 的API)提供了16 个变量条目，包括舞蹈性、节奏、响度和音调等音乐特征，以及针对98,340 首歌曲的artist\_name 和artist\_id。这些数据用于创建两个汇总数据集，包括:

a. 艺术家的平均值“data\_by\_artist”，

b. 表示跨年“data\_by\_year”。

|  |  |  |  |
| --- | --- | --- | --- |
|  | artist\_names | 表演曲目的艺术家。(数组) |  |
|  | artists\_id | impact \_ data . CSV 文件中给出的唯一标识号相同。（数字或者字符串） |  |
| 音  乐  的  特  征 | danceability | 根据音乐元素(包括节奏、节奏稳定性、节拍强度和整体规律性)的组合来衡量一条赛道是否适合跳舞的指标。 | 值为0.为最不适合跳舞，值为1.0 最适合跳舞。 |
| energy | 一种表示对强度和活动性的感知的量度。 | 值为0.0 表示最不强烈/充满活力，  值为1.0表示最强烈/充满活力。 |
| valence | 描述一首曲目所传达的音乐积极程度的一种衡量标准 | 值0.0 表示最负，  值1.0 表示最正 |
| tempo | 曲目的总体估计速度，以每分钟节拍(BPM)为单位。 |  |
| loudness | 音轨的整体响度，以分贝(DB)为单位。 | 值的典型范围在-60 到0 db 之间。 |
| mode | 曲目的模态(大调或小调)、旋律内容来源的音阶类型的指示 | 主要由1 表示，  次要为0。 |
| key | 估计的音轨的整体数值。为使用标准音调等级记法映射到音调的整数。 | 例如，0=C，1=C♯/D♭，2=D，依此类推。如果未检测到key 值，则的值设定为-1。(整数) |
| 人  声  类  型 | acousticness | 音轨是否是声学的置信度测量(没有技术增强或电子放大)。 | 值1.0 表示音轨的可信度很高。(浮点型) |
| instrumentalness | 预测曲目是否不包含人声。  在这种情况下，“噢”和“啊”的声音被视为器乐。  说唱或有声歌曲显然是“有声的”。 | 器乐度值越接近1.0，曲目中不包含声音内容的可能性就越大。高于0.5 的值用于表示乐器音轨，但随着该值接近1.0，置信度会更高。(浮点型) |
| liveness | 检测曲目中是否有观众。  活跃度值越高，表示实时执行曲目的概率越高。 | 如果值大于0.8，则表示音轨处于活动状态的可能性很大。(浮点数) |
| speechiness | 检测音轨中是否存在口语单词。 | 越是专门的演讲——比如录音(例如脱口秀、有声读物、诗歌)，属性值越接近1.0。大于0.66 的值描述可能完全由口语单词组成的曲目。  介于0.33 和0.66 之间的值描述可能同时包含音乐和语音的曲目(分段或分层)，包括说唱音乐等情况。低于0.33 的值很可能表示音乐和其他非语音轨道。(浮点数) |
| explicit | 检测曲目中的显式歌词 | (TRUE(1)=是，是；FALSE(0)=否，不是或未知)。**(**布尔值**)** |
| 描  述 | duration\_ms | 音轨的持续时间，以毫秒为单位。 | (整数) |
| popularity | 这条音轨的受欢迎程度。 | 该值将介于0 和100 之间，其中100 是最受欢迎的。 |
| year | 发行曲目的年份。(1921 年至2020 年的整数) |  |
| release\_date | 发行曲目的日期大多采用yyyy-mm-dd 的格式，但是日期的精度可能会有所不  同，有些只是以yyyy 给出。 |  |
| song\_title (censored) | 曲目的名称。 |  |

Data\_by\_year

100条数据

1921年到2020年 按年份升序排列

每年创作的音乐的特性

|  |  |
| --- | --- |
| year | 年份 |
| danceability | 根据音乐元素(包括节奏、节奏稳定性、节拍强度和整体规律性)的组合来衡量一条赛道是否适合跳舞的指标。 |
| energy | 一种表示对强度和活动性的感知的量度。是否强烈。 |
| valence | 描述一首曲目所传达的音乐积极程度的一种衡量标准 |
| Tempo | 曲目的总体估计速度，以每分钟节拍(BPM)为单位。 |
| loudness | 音轨的整体响度，以分贝(DB)为单位。 |
| Mode | 曲目的模态(大调或小调)、旋律内容来源的音阶类型的指示 |
| Key | 估计的音轨的整体数值。为使用标准音调等级记法映射到音调的整数。 |
| Acousticness | 音轨是否是声学的置信度测量(没有技术增强或电子放大)。 |
| instrumentalness | 预测曲目是否不包含人声。 |
| liveness | 检测曲目中是否有观众。  活跃度值越高，表示实时执行曲目的概率越高。 |
| speechiness | 检测音轨中是否存在口语单词。 |
| duration\_ms | 音轨的持续时间，以毫秒为单位。 |
| popularity | 这条音轨的受欢迎程度。 |

data\_by\_artist 5854条数据

5854位艺术家创作的音乐的特性。

该表格按count（每位艺术家参与创作的歌曲的数目）降序排列

|  |  |
| --- | --- |
| artist\_name | 艺术家姓名 |
| artist\_id | 艺术家id |
| danceability | 根据音乐元素(包括节奏、节奏稳定性、节拍强度和整体规律性)的组合来衡量一条赛道是否适合跳舞的指标。 |
| energy | 一种表示对强度和活动性的感知的量度。是否强烈。 |
| valence | 描述一首曲目所传达的音乐积极程度的一种衡量标准 |
| Tempo | 曲目的总体估计速度，以每分钟节拍(BPM)为单位。 |
| loudness | 音轨的整体响度，以分贝(DB)为单位。 |
| Mode | 曲目的模态(大调或小调)、旋律内容来源的音阶类型的指示 |
| Key | 估计的音轨的整体数值。为使用标准音调等级记法映射到音调的整数。 |
| Acousticness | 音轨是否是声学的置信度测量(没有技术增强或电子放大)。 |
| instrumentalness | 预测曲目是否不包含人声。 |
| liveness | 检测曲目中是否有观众。  活跃度值越高，表示实时执行曲目的概率越高。 |
| speechiness | 检测音轨中是否存在口语单词。 |
| duration\_ms | 音轨的持续时间，以毫秒为单位。 |
| popularity | 这条音轨的受欢迎程度。 |
| count | 每位艺术家创作的歌曲的数目  注：同一首歌曲可能有多个艺术家合作完成，所以该列求和大于98340首歌曲 |