问题 d: 音乐的影响

Music has been part of human societies since the beginning of time as an essential component of cultural heritage. As part of an effort to understand the role music has played in the collective human experience, we have been asked to develop a method to quantify musical evolution. There are many factors that can influence artists when they create a new piece of music, including their innate ingenuity, current social or political events, access to new instruments or tools, or other personal experiences. Our goal is to understand and measure the influence of previously produced music on new music and musical artists.

音乐自古以来就是人类社会的一部分，是文化遗产的重要组成部分。作为理解音乐在人类集体经历中所扮演角色的努力的一部分，我们被要求开发一种量化音乐进化的方法。有许多因素可以影响艺术家创造一个新的音乐作品，包括他们与生俱来的聪明才智，当前的社会或政治事件，获得新的乐器或工具，或其他个人经验。我们的目标是理解和衡量先前创作的音乐对新音乐和音乐艺术家的影响。

Some artists can list a dozen or more other artists who they say influenced their own musical work. It has also been suggested that influence can be measured by the degree of similarity between song characteristics, such as structure, rhythm, or lyrics. There are sometimes revolutionary shifts in music, offering new sounds or tempos, such as when a new genre emerges, or there is a reinvention of an existing genre (e.g. classical, pop/rock, jazz, etc.). This can be due to a sequence of small changes, a cooperative effort of artists, a series of influential artists, or a shift within society.

一些艺术家可以列出十几个或更多的其他艺术家，他们说谁影响了他们自己的音乐作品。也有人认为，影响力可以通过歌曲特征之间的相似程度来衡量，比如结构、节奏或歌词。音乐有时会发生革命性的转变，提供新的声音或节奏，例如当一种新的流派出现时，或者对现有流派(例如古典、流行/摇滚、爵士等)进行重新发明。这可能是由于一系列的小变化，艺术家的合作努力，一系列有影响力的艺术家，或社会内部的转变。

Many songs have similar sounds, and many artists have contributed to major shifts in a musical genre. Sometimes these shifts are due to one artist influencing another. Sometimes it is a change that emerges in response to external events (such as major world events or technological advances). By considering networks of songs and their musical characteristics, we can begin to capture the influence that musical artists have on each other. And, perhaps, we can also gain a better understanding of how music evolves through societies over time.

许多歌曲都有类似的声音，许多艺术家对音乐流派的重大转变做出了贡献。有时这些转变是由于一个艺术家影响另一个。有时候，这种变化是对外部事件(如重大世界事件或技术进步)的反应。通过考虑歌曲的网络及其音乐特征，我们可以开始捕捉音乐艺术家之间的相互影响。也许，我们还可以更好地理解音乐是如何随着时间的推移在社会中演变的。

Your team has been identified by the Integrative Collective Music (ICM) Society to develop a model that measures musical influence. This problem asks you to examine evolutionary and revolutionary trends of artists and genres. To do this, your team has been given several data sets by the ICM:

你的团队已经被整合集体音乐(ICM)协会确定为开发一个衡量音乐影响力的模型。这个问题要求你检查艺术家和流派的进化和革命趋势。为了做到这一点，ICM 给了你的团队几组数据:

1) “influence\_data” 1 represents musical influencers and followers, as reported by the artists themselves, as well as the opinions of industry experts. These data contains influencers and followers for 5,854 artists in the last 90 years.

“影响力数据”1代表音乐影响者和追随者，正如艺术家自己所报告的，以及行业专家的意见。这些数据包含了过去90年间5854位艺术家的影响者和追随者。

2) “full\_music\_data”2 provides 16 variable entries, including musical features such as danceability, tempo, loudness, and key, along with artist\_name and artist\_id for each of 98,340 songs. These data are used to create two summary data sets, including:

“完整的音乐数据”2提供了16个可变条目，包括音乐特征，如舞蹈性、节奏、响度和关键，以及艺术家名称和艺术家 id是我98,340首歌曲。这些数据用于创建两个汇总数据集，包括:

a. mean values by artist “data\_by\_artist”,

艺术家的平均值“艺术家的数据”

b. means across years “data\_by\_year”.

意思是跨年的“逐年的数据”。

Note: DATA provided in these files are a subset of larger data sets. These files CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.

注意: 这些文件中提供的数据是较大数据集的一个子集。这些文件包含唯一的数据，你应该使用这个问题。

To carry out this challenging project, the ICM Society asks your teams to explore the evolution of music through the influence across musical artists over time, by doing the following:

为了执行这个具有挑战性的项目，ICM 协会要求你的团队通过跨越音乐艺术家的时间影响来探索音乐的演变，具体做法如下:

1.Use full\_music\_data and/or the two summary data sets (with artists and years) of music

characteris

tics, to develop measures of music similarity. Using your measure, are artists

within genre more similar than artists between genres?

使用full\_music\_data和/或音乐特征的两个总结数据集(包含艺术家和年份)来开发音乐相似性度量。使用你的衡量标准，同一类型的艺术家是否比不同类型的艺术家更相似?

2.Compare similarities and influences between and within genres. What distinguishes a

genre and how do genres change over time? Are some genres related to others?

比较不同流派之间的相似之处和影响。流派的区别是什么?流派是如何随着时间变化的?有些流派与其他类流派关联吗?

3.Indicate whether the similarity data, as reported in the data\_influence data set, suggest

that the identified influencers in fact influence the respective artists. Do the ‘influencers’

actually affect the music created by the followers? Are some music characteristics more

‘contagious’ than others, or do they all have similar roles in influencing a particular

artist’s music?

说明data\_influence数据集中报告的相似度数据是否表明所识别的影响者实际上会影响各自的艺术家。这些“影响者”真的会影响追随者创作的音乐吗?是某些音乐特征比其他特征更具有“感染力”，还是它们在影响特定艺术家的音乐方面都有相似的作用?

4.Identify if there are characteristics that might signify revolutions (major leaps) in musical

evolution from these data? What artists represent revolutionaries (influencers of major

change) in your network?

从这些数据中确定是否有一些特征可能标志着音乐进化的革命(主要飞跃)?哪些艺术家代表了你网络中的革命者(重大变化的影响者)?

5.Analyze the influence processes of musical evolution that occurred over time in one

genre. Can your team identify indicators that reveal the dynamic influencers, and explain

how the genre(s) or artist(s) changed over time?

分析一种音乐流派随时间而发生的音乐演变的影响过程。你的团队能否识别出能够揭示动态影响因素的指标，并解释类型或艺术家是如何随着时间而改变的?

6.How does your work express information about cultural influence of music in time or

circumstances? Alternatively, how can the effects of social, political or technological

changes (such as the internet) be identified within the network?

你的作品如何表达音乐在时间或环境中的文化影响?另外，如何在网络内识别社会、政治或技术变革(如互联网)的影响?

7.Write a one-page document to the ICM Society about the value of using your approach to

understanding the influence of music through networks. Considering the two problem data sets were limited to only some genres, and subsequently to those artists common to both data sets, how would your work or solutions change with more or richer data? Recommend further study of music and its effect on culture.

向ICM协会写一份一页纸的文件，说明使用你的方法通过网络了解音乐影响的价值。考虑到这两个问题数据集仅局限于某些类型，随后又局限于这两个数据集所共有的艺术家，你的工作或解决方案将如何随着更多或更丰富的数据而改变?建议进一步研究音乐及其对文化的影响。

8.The ICM Society, an interdisciplinary and diverse group from the fields of music, history, social science, technology, and mathematics, looks forward to your final report.

ICM协会，一个来自音乐，历史，社会科学，技术和数学领域的跨学科和多样化的团体，期待你的最终报告。

Your PDF solution of no more than 25 total pages should include:

 One-page Summary Sheet.

 Table of Contents.

 Your complete solution.

 One-page document to ICM society.

 References list.

不超过25页的PDF解决方案应该包括:

一页摘要表。

目录。

您的完整解决方案。

给ICM协会的一页文件。

引用列表。

Note: New for 2021! The ICM Contest now has a 25-page limit. All aspects of your submission count toward the 25-page limit: Summary Sheet, Table of Contents, Main Body of Solution, Images and Tables, One-page Document, Reference List, and any Appendices.

注意:新的2021年!ICM竞赛现在有了25页的限制。你提交的所有方面都在25页的限制之内:摘要表，目录，解决方案的主体，图像和表格，一页的文档，参考列表，和任何附录。

Attachments

We provide the following four data files for this problem. THE DATA FILES PROVIDED

CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.

关于这个问题，我们提供了以下四个数据文件。所提供的数据文件包含您应该用于此问题的唯一数据。

1. influence\_data.csv

2. full\_music\_data.csv

3. data\_by\_artist.csv

4. data\_by\_year.csv

1. influence\_data.csv

影响数据

(Data is encoded in utf-8 to allow for handling of special characters):

(数据以 utf-8编码，以便处理特殊字符) :

- influencer\_id: A unique identification number given to the person listed as influencer. (string of digits)

Influencer \_ id: 给予被列为 influencer 的人的唯一识别号码。(一串数字)

- influencer\_name: The name of the influencing artist as given by the follower or industry experts. (string)

有影响力的艺术家的名字，由追随者或行业专家给出。(字符串)

- influencer\_main\_genre: The genre that best describes the bulk of the music produced by the influencing artist. (if available) (string)

Influencer \_ main \_ genre: 最能描述有影响力的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: The decade that the influencing artist began their music career. (integer)

有影响力的艺术家开始他们音乐生涯的十年

- follower\_id: A unique identification number given to the artist listed as follower. (string of digits)

Follower \_ id: 给予被列为 follower 的艺术家的唯一识别号码。(一串数字)

- follower\_name: The name of the artist following an influencing artist. (string)

Follower \_ name: 跟随一位有影响力的艺术家的艺术家的名字

- follower\_main\_genre: The genre that best describes the bulk of the music produced by the following artist. (if available) (string)

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)

- follower\_active\_start: The decade that the following artist began their music career. (integer)

追随者 \_ 活跃 \_ 开始: 以下艺术家开始他们音乐生涯的十年

2. full\_music\_data.csv 3. data\_by\_artist.csv 4. data\_by\_year.csv

2、完整的音乐数据。 csv 3、艺术家的数据4、年份的数据

Spotify audio features from the “full\_music\_data”, “data\_by\_artist”, “data\_by\_year”:

Spotify 音频功能来自“全音乐数据”、“艺术家数据”、“年份数据”:

- artist\_name: The artist who performed the track. (array)

艺术家姓名: 表演这首歌曲的艺术家。(数组)

- artist\_id: The same unique identification number given in the influence\_data.csv file. (string of digits)

Artist \_ id: 与 influence \_ data. csv 文件中给出的唯一标识号相同(数字串)

Characteristics of the music:

音乐特征:

- danceability: A measure of how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. (float)

舞蹈性: 根据音乐元素的组合，包括节奏、节奏稳定性、节拍强度和整体规律性，衡量一首歌曲适合跳舞的程度。0.0的值是最不适合跳舞的，1.0的值是最适合跳舞的。(浮动)

- energy: A measure representing a perception of intensity and activity. A value of 0.0 is least intense/energetic and 1.0 is most intense/energetic. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. (float)

能量: 表示对强度和活动的感知的量度。0.0是最不紧张/精力充沛的值，1.0是最紧张/精力充沛的值。一般来说，充满活力的音轨感觉快速、响亮和嘈杂。例如，死亡金属具有高能量，而巴赫前奏曲在音阶上得分较低。对这一属性有贡献的感知特征包括动态范围，感知响度，音色，发病率和一般熵。(浮动)

- valence: A measure describing the musical positiveness conveyed by a track. A value of 0.0 is most negative and 1.0 is most positive. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). (float)

配价: 描述一首歌曲所传达的音乐积极性的量度。0.0是最负面的，1.0是最正面的。价格高的音轨听起来更积极(例如快乐，愉快，愉快) ，而价格低的音轨听起来更消极(例如悲伤，沮丧，愤怒)。(浮动)

- tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. (float)

节奏: 以每分钟节拍为单位的一首歌曲的总体估计节奏(BPM)。在音乐术语中，节奏是给定乐曲的速度或节奏，直接来源于平均节拍持续时间。(浮动)

- loudness: The overall loudness of a track in decibels (dB). Values typical range between -60 and 0 db. Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). (float)

音量: 以分贝(dB)为单位的音轨的整体音量。值典型范围介于 -60和0分贝之间。响度值在整个轨道上取平均值，用于比较轨道的相对响度。响度是声音的质量，是身体强度(振幅)的主要心理相关因素。(浮动)

- mode: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track. Major is represented by 1 and minor is 0.

调式: 一首歌曲的调式(大调或小调) ，它的旋律内容来源于音阶的类型。Major 由1表示，小调为0。

- key: The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value for key is -1. (integer)

估计的整个轨道的关键。Integers 使用标准的 Pitch Class 符号映射到音高。例如0 = c，1 = c something/d something，2 = d，等等。如果没有检测到键，键的值是 -1。(整数)

Type of vocals:

声音类型:

- acousticness: A confidence measure of whether the track is acoustic (without technology enhancements or electrical amplification). A value of 1.0 represents high confidence the track is acoustic. (float)

声学性: 用来衡量音轨是否有声(没有技术增强或电子放大)的一种置信度。1.0的值表示高置信度的轨道是声学的。(浮动)

- instrumentalness: Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0. (float)

乐器性: 预测一首歌曲是否不包含人声。在这种情况下，“ Ooh”和“ aah”的声音被视为乐器。说唱或者口语音轨显然是“有声的”。乐器性值越接近1.0，这首歌曲就越有可能不包含声乐内容。大于0.5的值意在表示乐器音轨，但随着该值接近1.0，置信度会更高。(浮动)

- liveness: Detects the presence of an audience in a track. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live. (float)

活泼: 检测一首歌曲中是否有观众。更高的活力值表示这首歌被现场演奏的可能性增加。高于0.8的值提供了赛道实时的强烈可能性。(浮动)

- speechiness: Detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks. (float)

- 语言能力: 检测一个音轨中有没有口头语言。更专业的演讲-像录音(例如脱口秀，有声书，诗歌) ，更接近1.0的属性值。高于0.66的值描述了可能完全由口头语言构成的音轨。介于0.33和0.66之间的值描述了可能同时包含音乐和语音的曲目，可以是分段的，也可以是分层的，包括说唱音乐这样的情况。小于0.33的值最有可能代表音乐和其他非语音类曲目。(浮动)

- explicit: Detects explicit lyrics in a track (true (1) = yes it does; false (0) = no it does not OR unknown). (Boolean)

显式的: 检测一首歌曲中显式的歌词(true (1) = yes it does; false (0) = no it does not OR unknown)。(Boolean)

Description:

描述:

- duration\_ms: The duration of the track in milliseconds. (integer)

Duration \_ ms: 以毫秒为单位的音轨持续时间。(整数)

- popularity: The popularity of the track. The value will be between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played more frequently now will have a higher popularity than songs that were played more frequently in the past. Duplicate tracks (e.g. the same track from a single and an album) are rated independently. Artist and album popularity are derived mathematically from track popularity. (integer)

流行度: 这首歌的流行程度。这个数值在0到100之间，100是最受欢迎的。受欢迎程度是由算法计算出来的，而且在很大程度上是基于赛道的总播放次数以及这些播放次数的最新程度。一般来说，现在经常播放的歌曲会比过去经常播放的歌曲更受欢迎。重复的歌曲(例如单曲和专辑中的相同歌曲)是独立评分的。艺术家和专辑的受欢迎程度从数学上来源于曲目的受欢迎程度。(整数)

- year: The year of release of a track. (integer from 1921 to 2020)

年: 一首歌曲发行的年份。(1921年至2020年的整数)

- release\_date: The calendar date of release of a track mostly in yyyy-mm-dd format, however precision of date may vary and some just given as yyyy.

发行日期: 一首歌曲发行的日历日期，大部分是 yyyy-mm-dd 格式，不过日期的精确度可能会有所不同，有些只是给出 yyyy。

- song\_title (censored): The name of the track. (string) Software was run to remove any potential explicit words in the song title.

歌曲名称(审查) : 歌曲的名称。(字符串)软件运行删除任何潜在的明确的话在歌曲标题。

- count: The number of songs a particular artist is represented in the full\_music\_data.csv file. (integer)

Count: 在完整的 \_ music \_ data. csv 文件中表示特定歌手的歌曲数量