

Spotify 数据分析

Spotify 是瑞典的一家流媒体音乐服务平台。我们对其在网上开放的音乐数据进行处理与分析。

数据集: tracks.csv （注意：不要用 excel 软件打开，否则会造成乱码。）

该数据集中存放的是歌曲的各种属性信息，每一行对应的是一个数据项，包括但不限于名称、歌手、时长等信息。。

作业内容：

1. 将 tracks 中数据加载到一个 dataframe 对象中，并查看其全局信息。
2. 对 tracks 中的数据的信息进行展示，要求每一行是数据名称，每一列为统计的类别名称。
3. 分别找出最流行和最不流行的 10 首歌。
4. 将歌曲的发行时间 (release_date) 设置成索引,并将其设置成日期类型。要求：直接对 tracks 进行修改，而不是重新创建一个 dataframe 对象。
5. 查看最晚发行和最早发行的 5 首歌。
6. 查看 'artists' 列第 18 行中的数据。
7. 将 'duration_ms' 列中的单位从毫秒 (millisecond) 转换为秒 (second), 并将该列名称改为 'duration'。
8. 选取 100 条数据项，并将 'danceability', 'energy', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo' 列的数据分别展示在 8 个子图 (2 行 4 列，散点图) 中。要求：每一个子图中点的颜色不一样；并给每个子图设置标题。画完图之后，通过分析图像尝试总结歌曲的特征。
9. 选取 200 条数据项，利用 pandas 中的 scatter_matrix 函数，将 ['danceability', 'speechiness', 'tempo', 'valence'] 中特征之间的分布关系展示出来。通过分析图像尝试总结特征之间的关联关系。

作业要求：

1. 代码
2. 数据分析和说明

附：

属性说明：

1. **Acousticness** (原声性，非电子音程度) : describes how acoustic the track is
2. **Danceability**: (舞曲) represents how suitable a track is for dancing, a combination of other elements like tempo and beat strength
3. **Energy** (强度) : represents a perceptual measure of intensity (metal is likely to have high energy and classical likely to have low energy)
4. **Instrumentalness** (歌唱部分占比) : predicts whether a track contains vocals ('ooh' and 'aah' sounds are treated as instrumental)
5. **Loudness** (响度) : describes the overall average loudness of a track
6. **Speechiness** (朗诵比例/说唱) : detects the presence of spoken word (rap music is likely to have a higher speechiness)
7. **Tempo** (节奏、速度) : overall estimated pace of the track in beats per minute (BPM)
8. **Valence** (心理感受) : describes the positiveness of a track (tracks with high valence sound happy, and those with low valence sound sad)
9. **Liveness** (现场感)