

	状态	举办方	赛季2	奖金	参赛队伍
阿里音乐流行趋势预测大赛	已结束		2016-07-15	¥ 380000	5475

- 赛制
- 赛题与数据
- 排行榜
- 论坛
- FAQ

文档 (注册之后可以下载)	格式
(*****t.CSV	.(368B) <a href="#">↓</a>
p*****s.CSV	.(2MB) <a href="#">↓</a>
p*****s.zip	.(315MB) <a href="#">↓</a>
p*****U3l4)	.(.) <a href="#">↓</a>

竞赛题目

经过7年的发展与沉淀，目前阿里音乐拥有数百万的曲库资源，每天千万的用户活跃在平台上，拥有数亿人次的用户试听、收藏等行为。在原创艺人和作品方面，更是拥有数万的独立音乐人，每月上传上万个原创作品，形成超过几十万首曲目的原创作品库，如此庞大的数据资源库对于音乐流行趋势的把握有着极为重要的指引作用。

本次大赛以阿里音乐用户的历史播放数据为基础，期望参赛者可以通过对阿里音乐平台上每个阶段艺人的试听量的预测，挖掘出即将成为潮流的艺人，从而实现对一个时间段内音乐流行趋势的准确把控。

竞赛数据

(注：每个赛季换数据阶段，表名加前缀” p2\_ ”，如p2\_mars\_tianchi\_user\_actions)

大赛开放抽样的歌曲艺人数据，以及和这些艺人相关的6个月内（20150301-20150830）的用户行为历史记录。

用户行为表（mars\_tianchi\_user\_actions）

列名	类型	说明	示例
user_id	String	用户唯一标识	7063b3d0c075a4d276c5f06f4327cf4a
song_id	String	歌曲唯一标识	effb071415be51f11e845884e67c0f8c
gmt_create	String	用户播放时间（unix时间戳表示）精确到小时	1426406400
action_type	String	行为类型：1，播放；2，下载，3，收藏	1
Ds	String	记录收集日（分区）	20150315

注：用户对歌曲的任意行为为一行数据。

歌曲艺人（mars\_tianchi\_songs）

列名	类型	说明	示例
song_id	String	歌曲唯一标识	c81f89cf7edd24930641afa2e411b09c
artist_id	String	歌曲所属的艺人Id	03c6699ea836decbc5c8fc2dbae7bd3b

publish_time	String	歌曲发行时间，精确到天	20150325
song_init_plays	String	歌曲的初始播放数，表明该歌曲的初始热度	0
Language	String	数字表示1,2,3...	100
Gender	String	1,2,3	1

结果集：  
选手需要预测艺人随后2个月，即60天（20150901-20151030）的播放数据。

选手提交结果表（mars\_tianchi\_artist\_plays\_predict）

列名	类型	说明	示例
artist_id	String	歌曲所属的艺人Id	023406156015ef87f99521f3b343f71f
Plays	String	艺人当天的播放数据	5000
Ds	String	日期	20150901

选手需要预测9月1日至10月30日60天内所有艺人的结果。

- 初赛，选手通过天池网站提交结果（临近评测开始时间开放，左侧会多一个“提交结果”导航），文件命名方式：  
mars\_tianchi\_artist\_plays\_predict.csv；字段之间以逗号分隔；编码格式为UTF-8。结果格式如下图：

```
023406156015ef87f99521f3b343f71f,165,20150902
023406156015ef87f99521f3b343f71f,112,20150903
023406156015ef87f99521f3b343f71f,193,20150904
023406156015ef87f99521f3b343f71f,114,20150905
023406156015ef87f99521f3b343f71f,117,20150906
023406156015ef87f99521f3b343f71f,123,20150907
023406156015ef87f99521f3b343f71f,128,20150908
```

- 复赛，选手直接在本队伍所在项目空间产出结果mars\_tianchi\_artist\_plays\_predict即可，表结构（字段名，字段类型，字段顺序）需与上述所列表结构保持一致。系统会自动根据表名去扫描和评测。注意：在数加平台读取表时，请在表面前加前缀odps\_tc\_257100\_f673506e024，形如：desc odps\_tc\_257100\_f673506e024.mars\_tianchi\_user\_actions

评估指标

设艺人j在第k天的实际播放数为 $T_{j,k}$ ，参赛选手集合为 $U$ ，艺人集合为 $W$ ，参赛选手i的程序计算得到艺人j在第k天的播放数为 $S_{i,j,k}$ ，则参赛选手i对艺人j的播放预测和实际的方差归一化方差 $\sigma_{i,j}$ 为：

$$\sigma_{i,j} = \sqrt{\frac{1}{N} \sum_{k=1}^N ((S_{i,j,k} - T_{j,k}) / (T_{j,k}))^2}$$

而艺人j在的权重根据艺人的播放量平方根：

$$\phi_j = \sqrt{\sum_{k=1}^N T_{j,k}}$$

参赛选手i的预测为 $F_i$

$$F_i = \sum_{j \in W} (1 - \sigma_{i,j}) * \phi_j$$

最终排名按照F值评判，F值越大，代表结果越优，排名越靠前。