

[首页](#) > [天池大赛](#) > [赛题与数据](#)



	状态	举办方	第 1 赛季截止日期	总奖池	参赛队
阿里音乐流行趋势预测大赛	进行中		2016/06/14	¥ 380000	99


  
 报名参赛

[赛制介绍](#)

[赛题与数据](#)

[FAQ](#)

[论坛](#)

## 赛题与数据

文件名称 (报名后可下载)	文件格式
mars_tianchi_songs.csv	.csv (889KB) <a href="#">↓</a>
mars_tianchi_user_actions.zip	.zip (111MB) <a href="#">↓</a>

### 竞赛题目

本次大赛以虾米音乐用户的历史播放数据为基础，期望参赛者可以通过对虾米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用户行为6个月（20150301-20150831）的历史记录。

#### 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：

需要选手预测艺人随后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

选手需预测**9月1日至10月30日60天**，所有艺人的结果。

第一赛季，选手通过天池网站提交结果， 文件命名mars\_tianchi\_artist\_plays\_predict.csv。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
```

第二赛季，选手直接在平台上本队伍所在项目产出结果mars\_tianchi\_artist\_plays\_predict即可（表结构（字段名，字段类型，字段顺序）需保持一致。

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的播放预测的归一化均方差 $\sigma_{ij}$ 为：

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

而艺人j的权重根据艺人的播放量等大小单调递增（具体公式不公布）

$$\phi_j = F(S)$$

参赛选手i的最终得分

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

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。