数据规格说明文档

原数据

movies.csv

编码格式

输入数据和输出数据均为 utf-8，且仅支持 utf-8

输入数据：

输入数据要求为一个单一文件，每一行对应一部电影，每行的数据格式如下：

百度索引指数，百度新闻指数，海报路径，类别；类别；… ，电影产地，总时长，上映年份，想要观看本电影的人数，电影名称

其中：

- ✧ 百度索引指数：

  上映前 31 天的索引指数，必须为 31 个数字，中间用逗号隔开，必须按照时间顺序排列

- ✧ 百度新闻指数

  同上，注意时间与百度索引指数一 一对应

- ✧ 海报路径

海报图片的相对路径或者绝对路径

- ✧ 类别

  电影的类别划分，每个类别用分号隔开，可能的类别如下：动画，儿童，战争，惊悚，科幻，音乐，家庭，武侠，传记，爱情，悬疑，运动，冒险，犯罪，历史，恐怖，灾难，动作，古装，奇幻，西部，喜剧，剧情，歌舞

- ✧ 电影产地

  可选的选项如下：进口，国产，进口特种，进口分账，进口报关

- ✧ 总时长

  单位为分钟

- ✧ 上映年份

  必须为四位数字

- ✧ 想要观看本电影的人数

  必须为纯数字，单位为人

- ✧ 电影名称

  可给出任意值，用于输出文件中，详见输出文件规格说明

示例：

30000,54000,51333,47333,90000,74000,70666,46000,44000,40000,40666,48666,62666,58000,46000,41333,42666,42666,48000,56000,52666,44000,45333,45333,46666,45333,53333,50666,46000,98666,157333,0,1,1,2,7,3,1,3,2,3,2,8,4,1,1,

6,11,7,3,2,2,5,6,9,7,14,7,6,15,14,18,imgs/p2217523448.jpg,

喜剧;动画;奇幻,国产,100,2014,7760,十万个冷笑话,

输出数据：

输出数据为一个单一文件，每一行一部电影，顺序与输入文件相对应，格式如下：

电影名称，电影票房

其中：

- 电影名称

  与输入数据中的电影名称相对应

- 电影票房

  单位万元

示例：

十万个冷笑话,13370