## 基于影视量化的票房分析

撰写时间：2018年3月10日

##### 数据收集情况

已收集到的相关数据有：

1. 2011-01-01至今（跟随网站动态更新）的每日票房总数、每部电影每日排片占比、上座率情况、发行时间等数据。

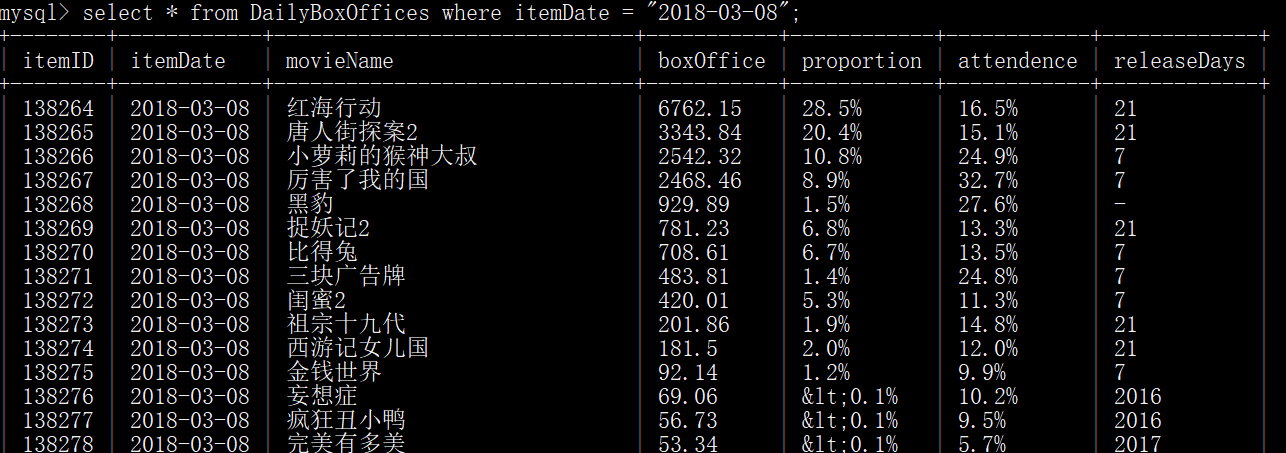


图1-1 每日票房数据示意图

1. 中国票房网站的所有数据，包含：电影名称、上映地点、影片时长、总票房、导演名称、电影类型、制作技术、上映日期、主要发行公司、主要制作公司、主要演员列表。
2. 已同步实时票房数据库获取每日的排片占比、票房信息、购票指数等。

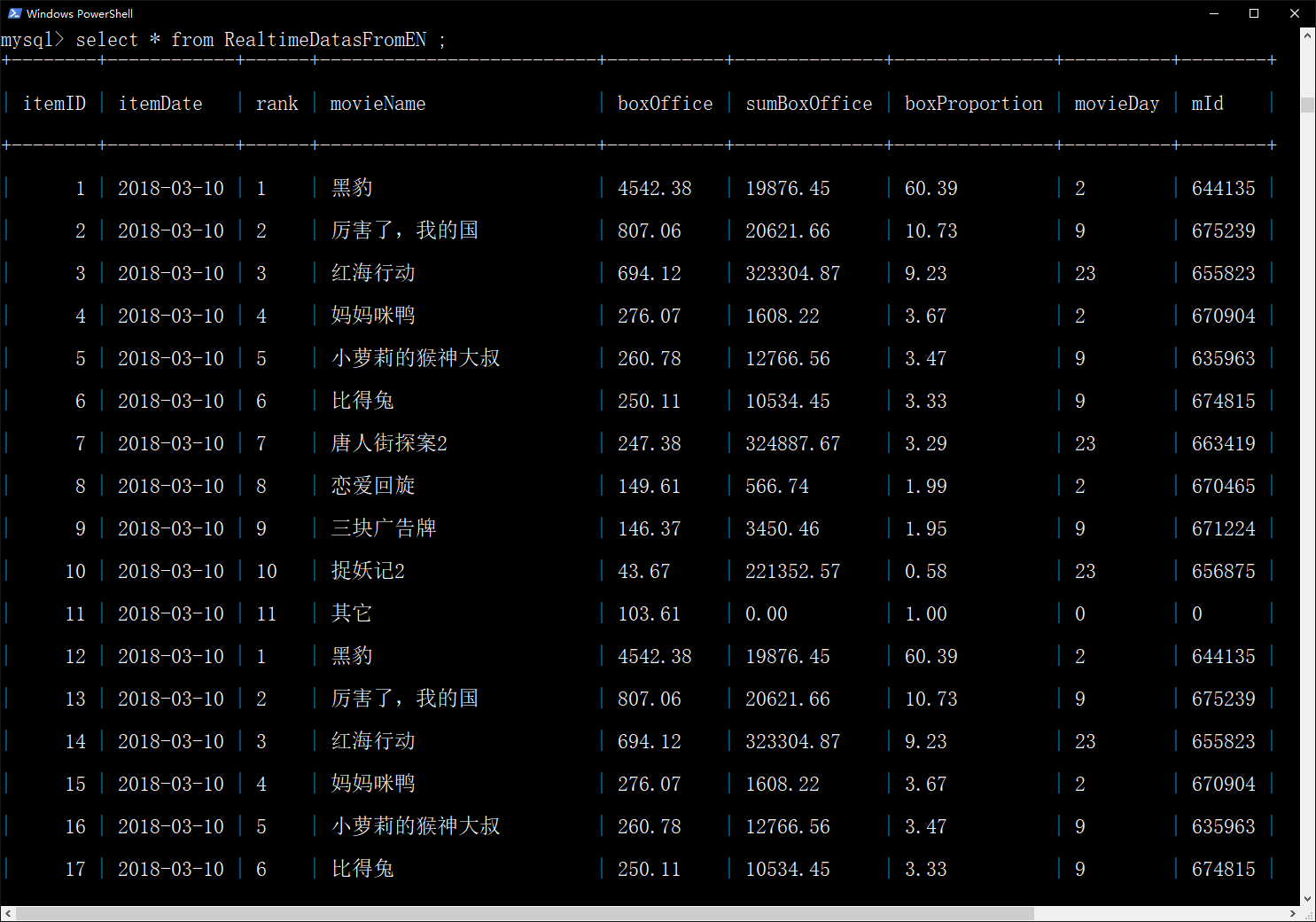


图1-2 实时票数据（自动每日更新）

同另一组协商后，12指标中还未获取的数据有：

演员的详细信息、编剧详细信息、点映信息、IP知识产权信息

##### 2. 数据库和代码情况

1. 关于代码，已建立GitHub代码仓库：

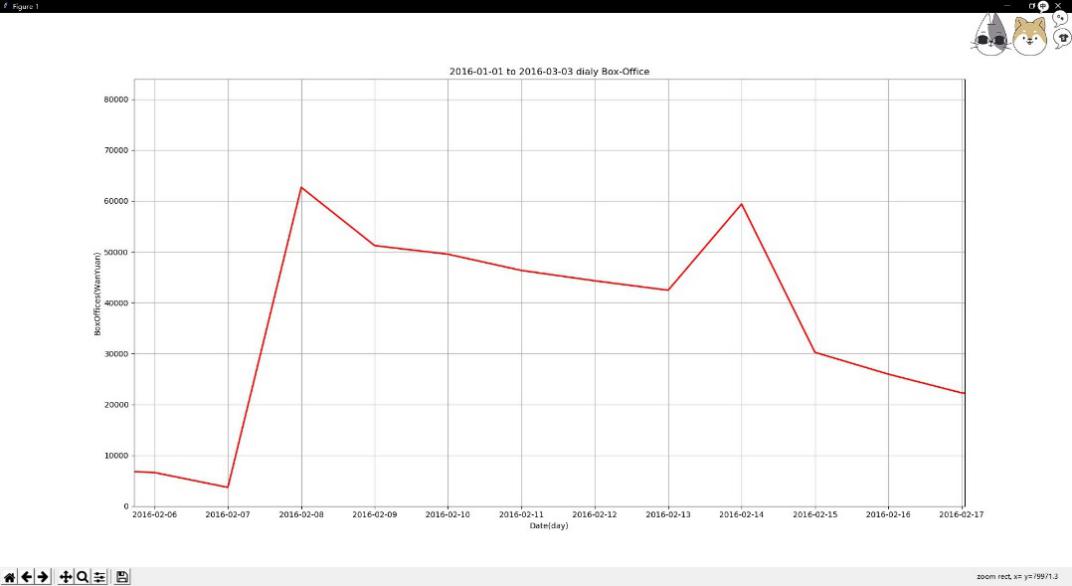
<https://github.com/zhaotianxiang/MovieBoxOfficeAnalysisGit>

1. 关于数据库，已在阿里云服务器中建立数据仓库，并且能够每日定时自动更新数据库的所有数据。数据库名称：movie ，远程连接方式：
2. mysql -h 47.100.51.19 -u root -p
3. password: aini1314@xiaoqing

##### 3. 数据量化和分析情况

1.已完成电影档期的量化分析

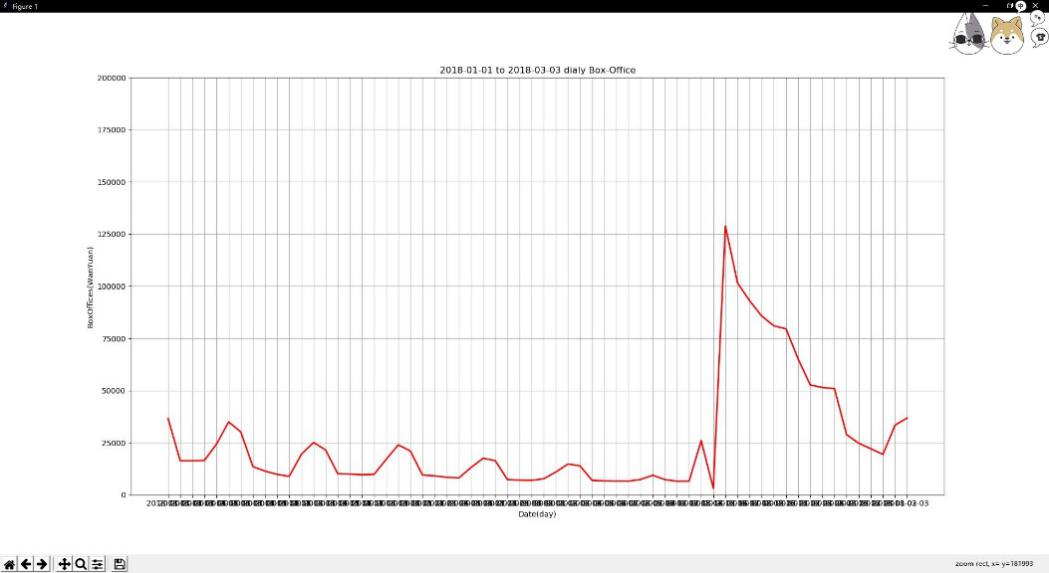
首先我们确定划分的档期有：元旦档、春节档、元宵档、清明档、五一档、端午档、十一档、情人节档、暑假档、寒假档。



**春节**

**除夕**

图3-1 2016年春节前后票房的明显变化



2018年春节

图3-2 2018年春节前后票房的明显变化

使用类似的每日票房线性变化的分析图示对每个档期进行简单的分析，在此不一一列举已将分析程序设计成简单接口，随时可以对数据库中任意时间段内的每日票房进行简单的画图分析。

根据日期进行档期的分类，然后计算票房的平均值进行量化，量化结果如下：

表3-1 档期量化后表

|  |  |
| --- | --- |
| **档期名称** | **单日平均票房/万元** |
| **春节** | 37904 |
| **情人节** | 19454 |
| **十一** | 16712 |
| **元旦** | 15494 |
| **中秋节** | 14273 |
| **元宵节** | 14033 |
| **端午节** | 13636 |
| **五一** | 12883 |
| **寒假** | 11558 |
| **暑假** | 11215 |
| **清明节** | 9763 |
| **七年单日平均值** | 9553 |

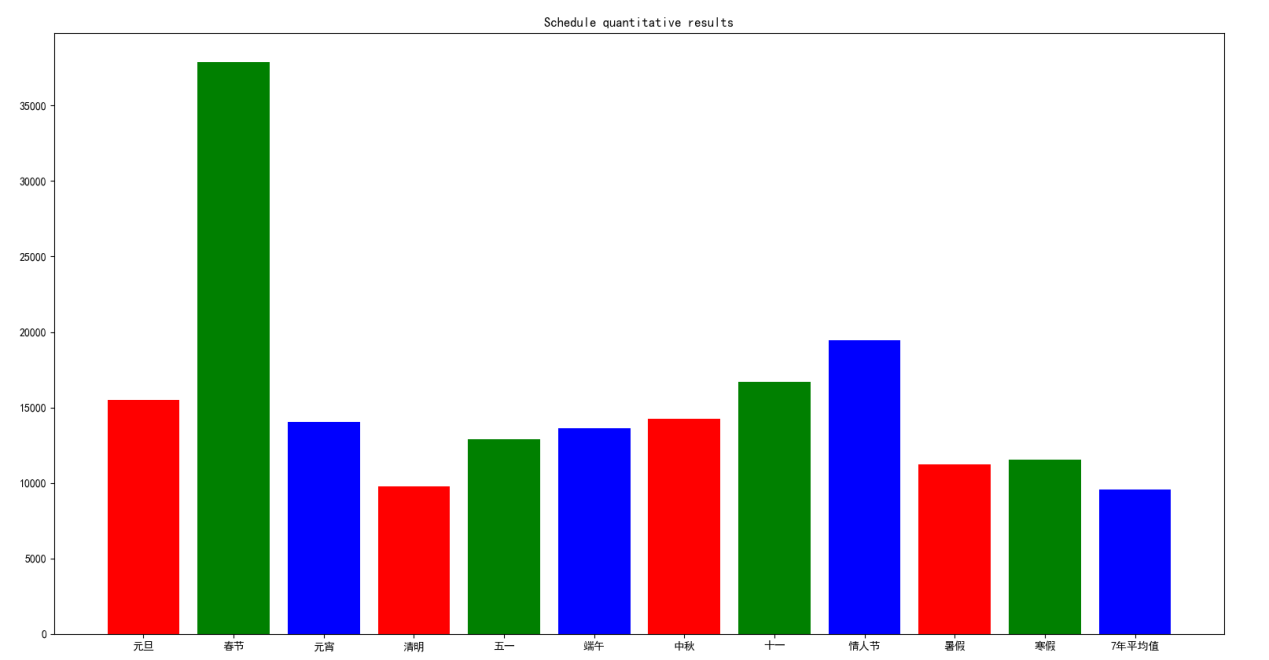


图3-3 档期量化柱状示意图