

(二) 电影数据分析

实战训练项目

项目简介：

对电影数据的分析，我们对今年所有的电影数据进行分析，并且可视化

2.大概分下面几个步骤:

- 1).获取数据
- 2).清洗数据
- 3).分析数据
- 4).可视化数据

3.具体的建议

- 1).电影的数据可以用tushare库去获取(不用去爬猫眼了)
- 2).数据的清洗会用Pandas
- 3).分析会用简单的统计学知识
- 4).我们统一用pyecharts

4.输出

要求通过数据分析，找出数据里面的秘密，你觉得最好的电影，从那个维度来分析并用可视化进行说明

1.环境配置

system : deepin

ide : vscode

python : anaconda 3

2.获取数据

使用tushare库

<https://tushare.pro>

注意：抓取电影数据要500个积分，但可以使用第三方短信接收工具（现在邀请注册有限制），我使用了易码平台

（<http://51ym.me/User/Payment.aspx>），0.1一条，最低充值10块。

下载最新版tushare SDK

方式1：

```
1. pip install tushare
```

如果安装网络超时可尝试国内pip源，如

```
1. pip install tushare -i https://pypi.tuna.tsinghua.edu.cn/simple
```

方式2：

访问<https://pypi.python.org/pypi/tushare/>下载安装
方式3：

访问<https://github.com/waditu/tushare>,将项目下载或者clone
到本地，进入到项目的目录下，执行：python setup.py install
导入tushare

```
1. import tushare as ts
```

设置token

```
1. ts.set_token('your token here')
```

初始化pro接口

```
1. pro = ts.pro_api()
```

如果上一步骤ts.set_token('your token')无效或不想保存token
到本地，也可以在初始化接口里直接设置token:

```
1. pro = ts.pro_api('your token')
```

思路：

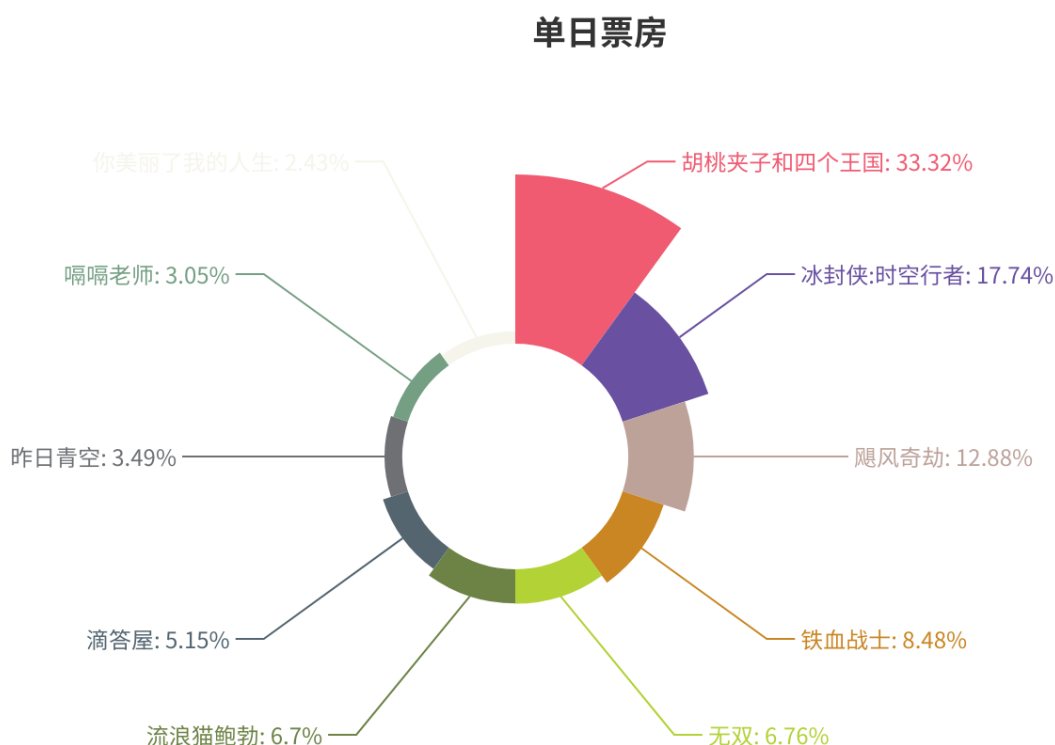
获取单日数据

放到mysql

可以保持更新数据

单日票房统计

利用当日票房参数统计



全年票房统计

从一月份开始判断

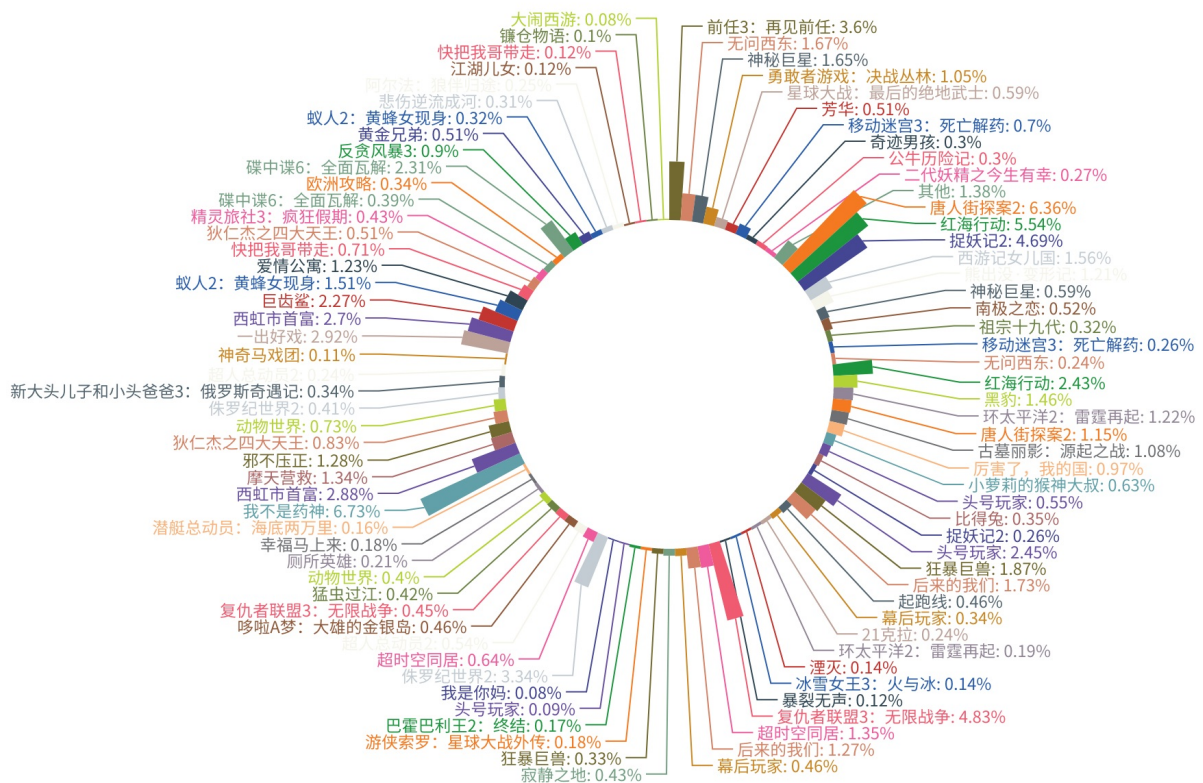
利用for if

if判断是否同一个电影名字，如果相同，两者票房相加

第二个if判断是否新增加电影，如果是，添加到一月数据里面

安装pyecharts可视化库

```
1. pip install pyecharts
```



参考

http://pyecharts.org/#/zh-cn/charts_base

<https://juejin.im/entry/5ac099626fb9a028de44d6eb>

https://blog.csdn.net/Xw_Classmate/article/details/51333646