

DATA MINING AND ANALYSIS OF NETEASE CLOUD MUSIC

网易云音乐的数据挖掘和分析

★ 汇报人：杜涵悦 ★

★ 学号：10181900144 ★

目录

C O N T E N T S

01 项目背景

02 研究工具与项目难点

03 数据采集

04 数据存储与清洗

05 数据分析及可视化



01

项目背景

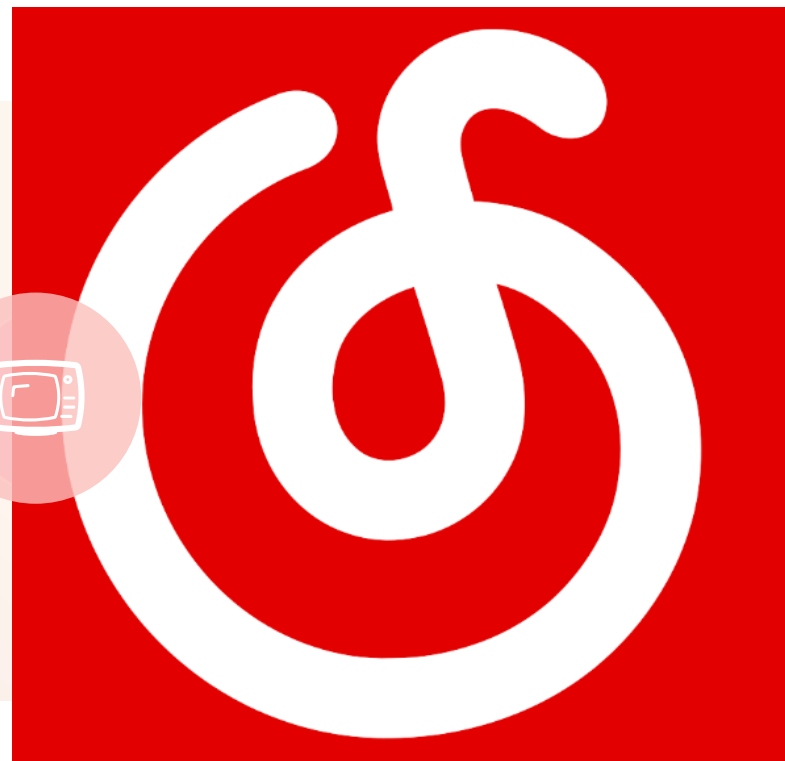


项目背景



兴趣所起

国事所致





02

研 究 工 具
与
项 目 难 点

爬取数据

如何解密?!?

网易云反爬机制!!
采用post获取数据反爬机制更严格!

难点

- 选取合适的线程，巧妙利用sleep。
- 使用代理服务器，时常清理cookie并更新headers参数

工具

工具

数据获取与存储分析

环境: Windows
python3.7
MySQL

数据可视化

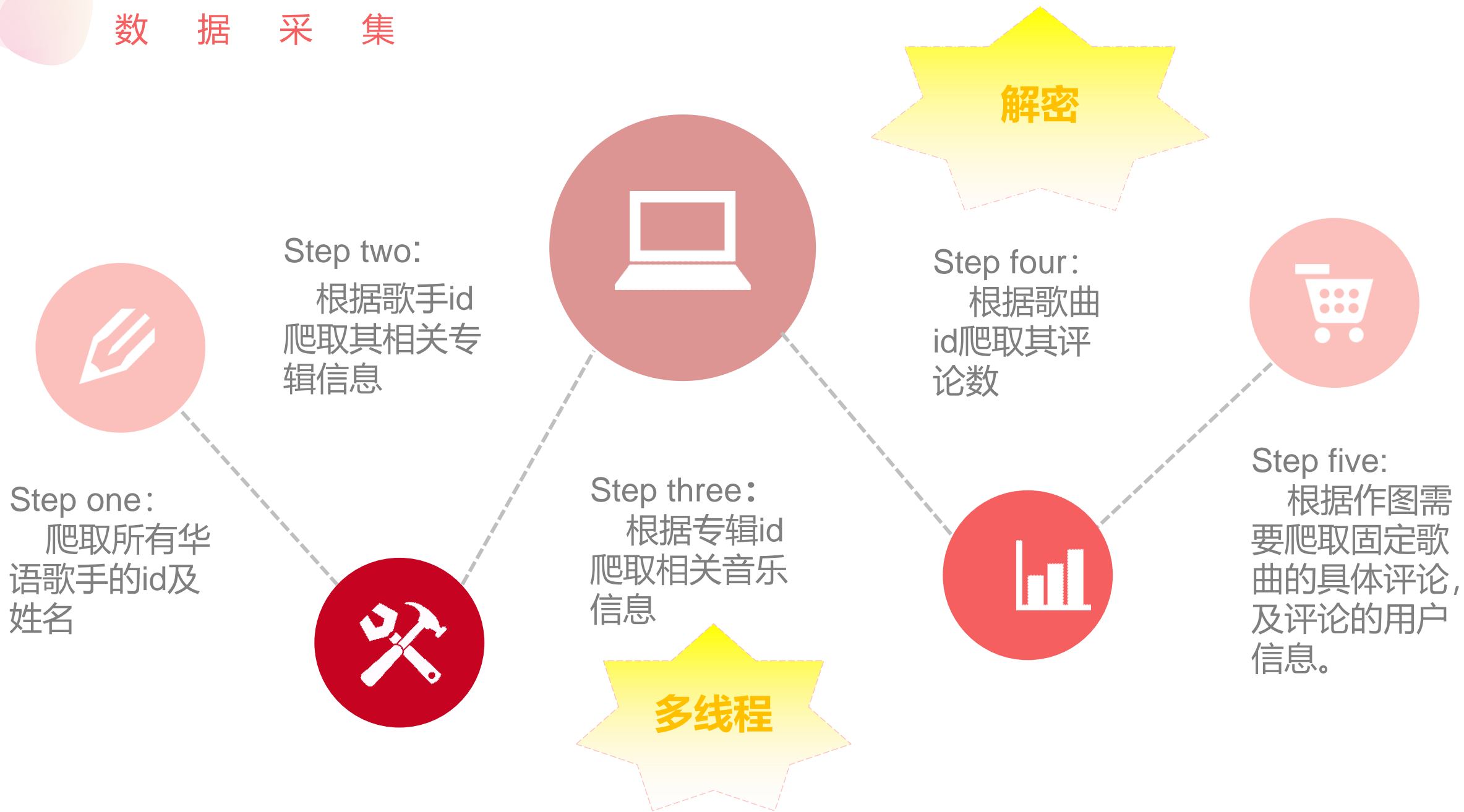
Pyecharts
matplotlib
Excel



03

数 据 采 集

数据采集



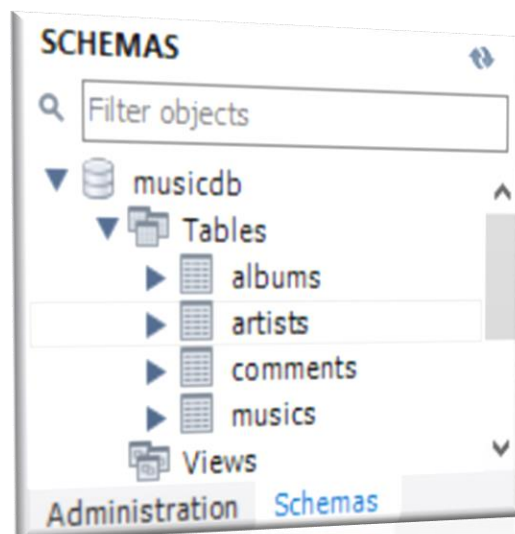


04

数 据 存 储
与
清 洗

数据存储与清洗

新建数据库及
table存放各
类数据



清除重复数据

```
create table musicdb.albums (select  
distinct * from musicdb.albums_old);
```

根据后期绘图需求筛选
评论数top100的音乐

```
select * from (select * from musicdb.comments  
order by COMMENTS desc) as tmp limit 100;
```





05

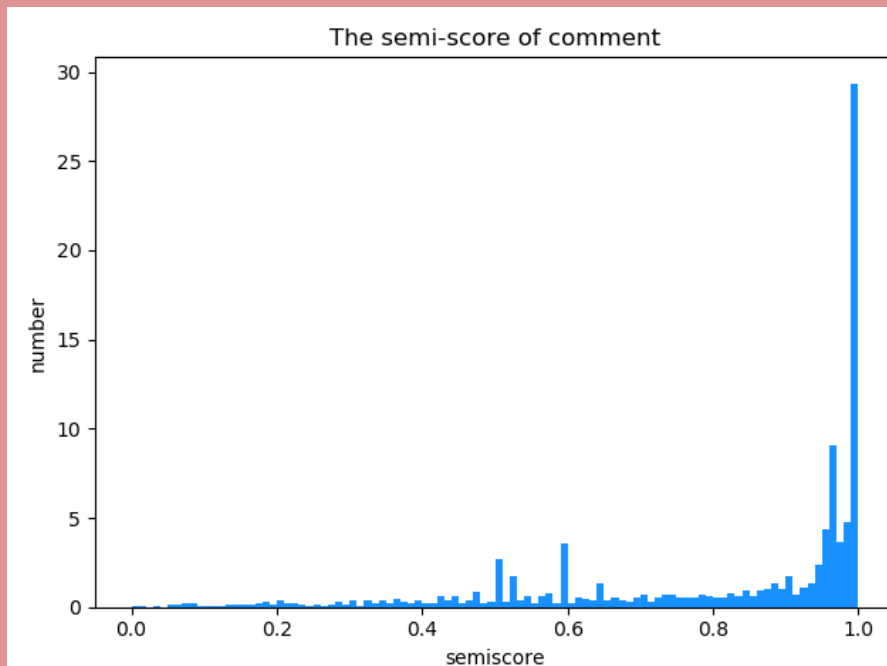
数 据 分 析
及 视 化
可

词云

VS

情感分析

生日
快乐



词云

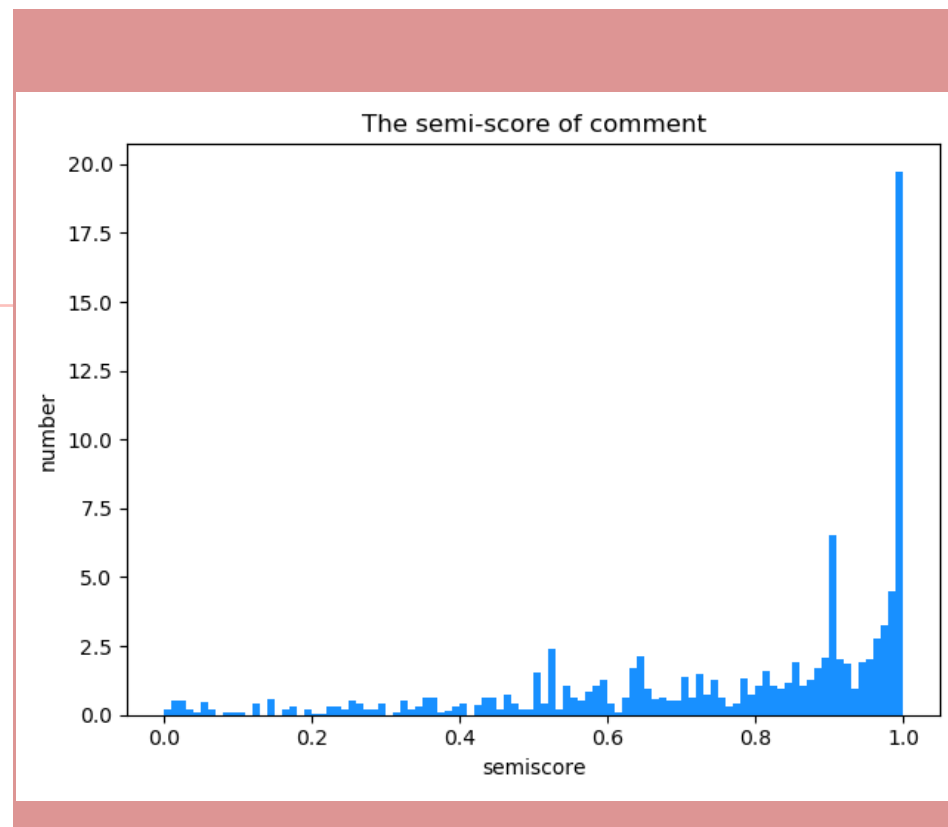
词云

VS

情感分析



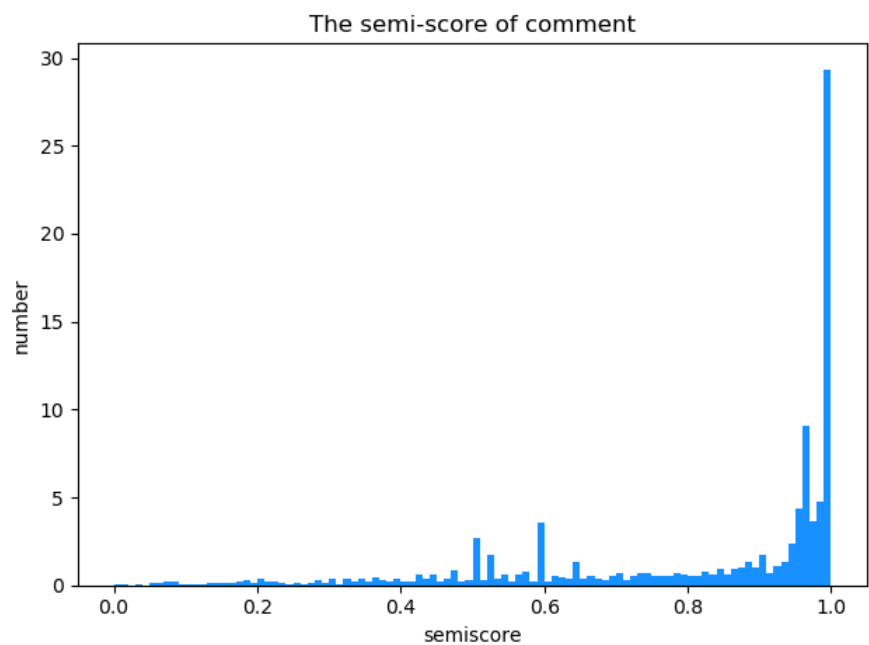
城南
花已
开



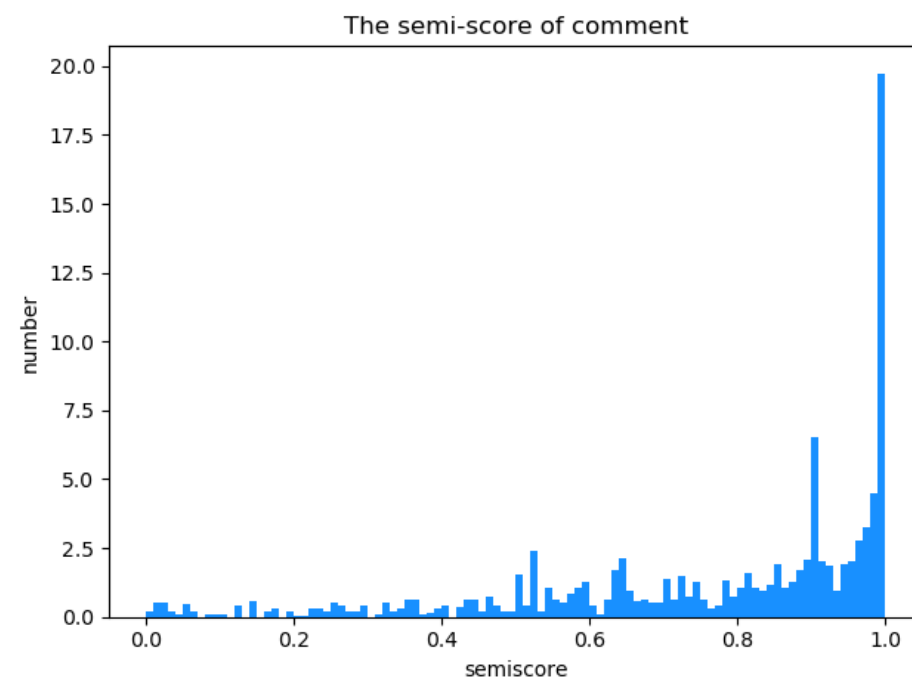
生日快乐

VS

城南花已开

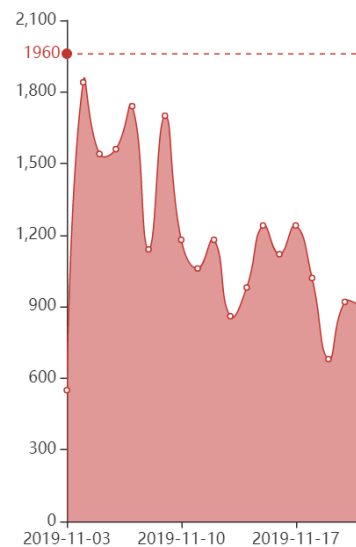


?



我和我的祖国 评论时间分布

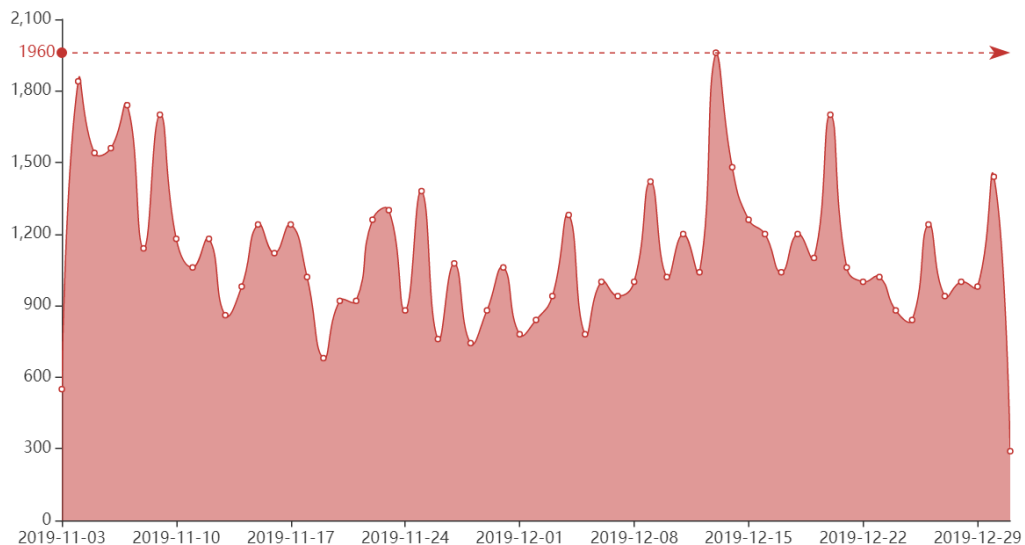
评论时间分布曲线（我和我的祖国）



我和我的祖国 评论时间分布曲线

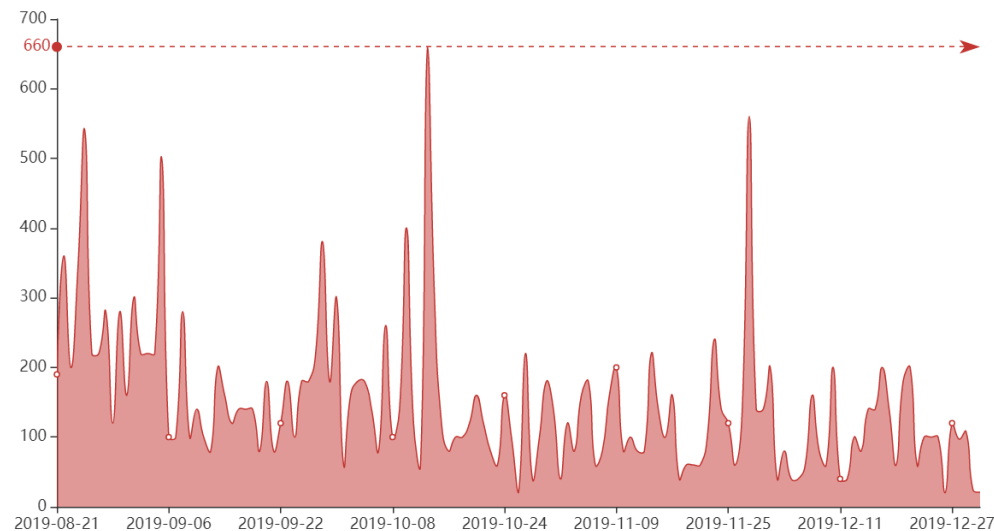
评论时间分布曲线（我和我的祖国）

—○— y



评论时间分布曲线（无所求必满载而归）

—○— y

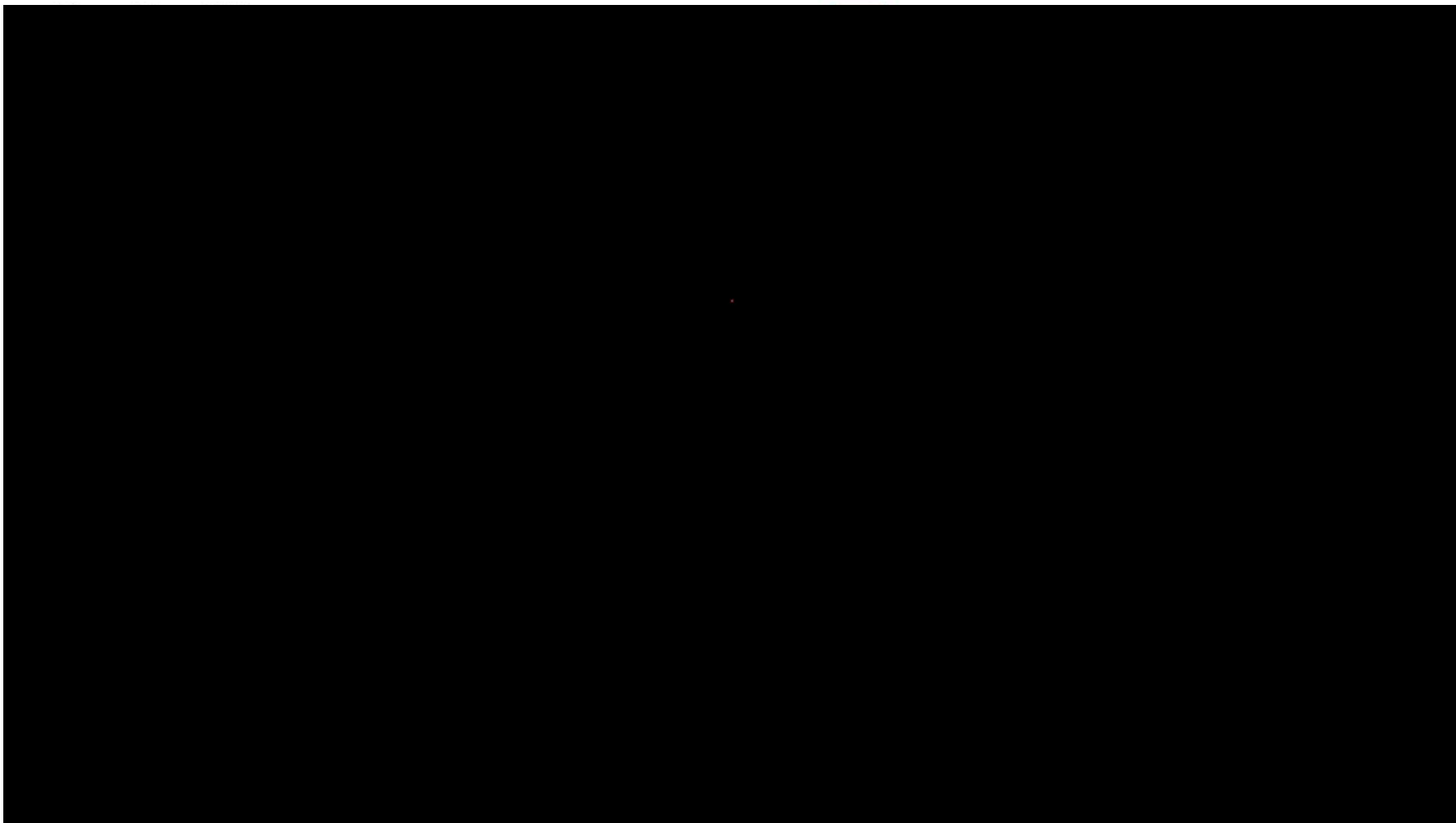


无所求必满载而归 评论时间分布曲线

数据分析及可视化



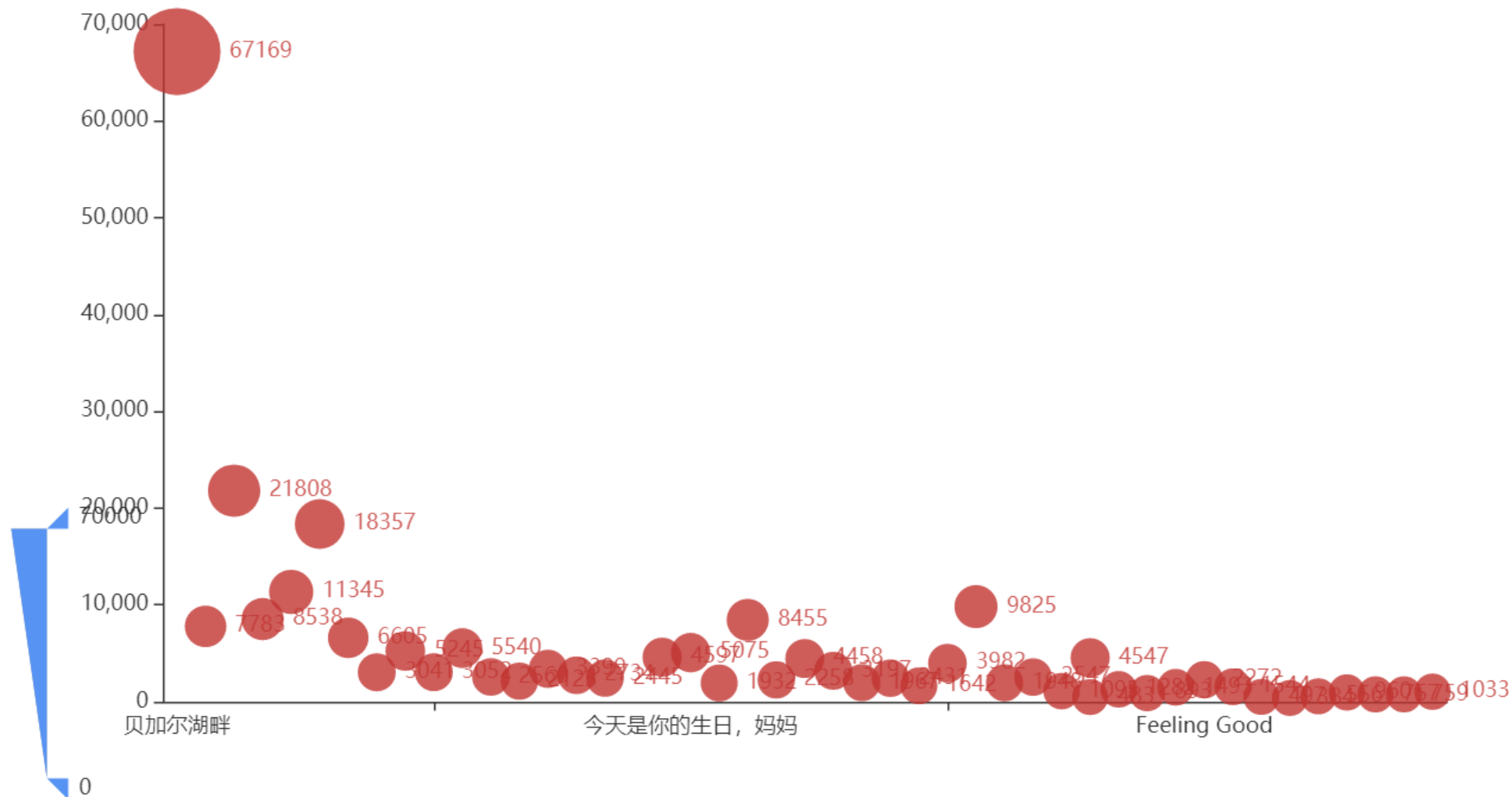
对评论数top100的音乐评论用户所在地区的密集度分析



数据分析及可视化

Scatter-VisualMap(Size)

● 评论数



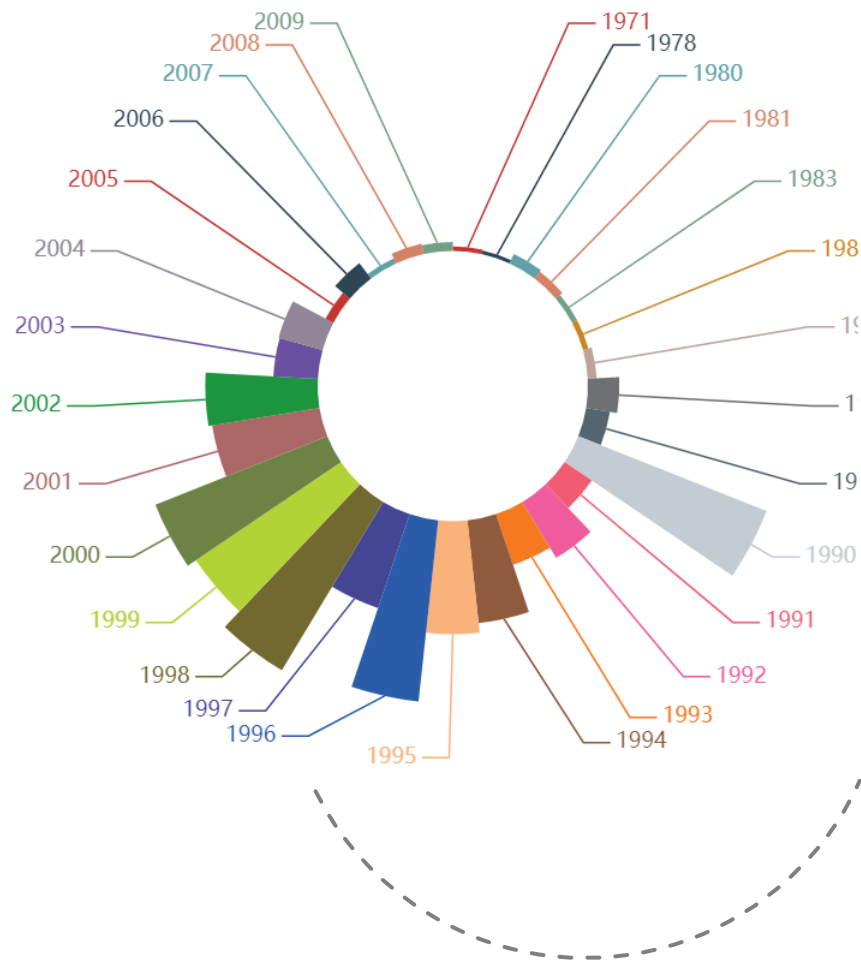
李健歌曲评论数



李健歌曲评论数 (除去贝加尔湖畔)

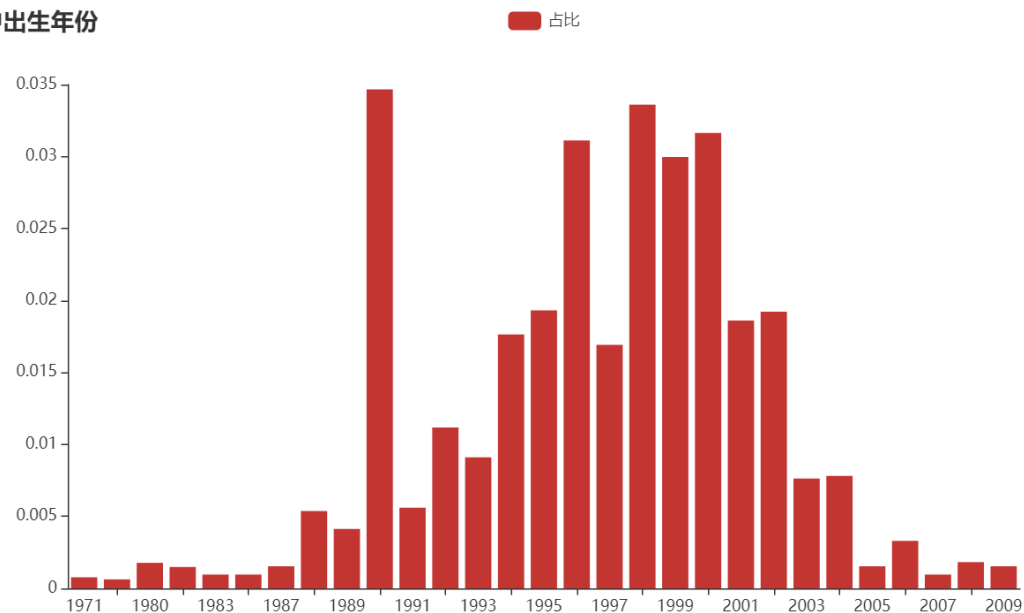
033

数据分析及可视化

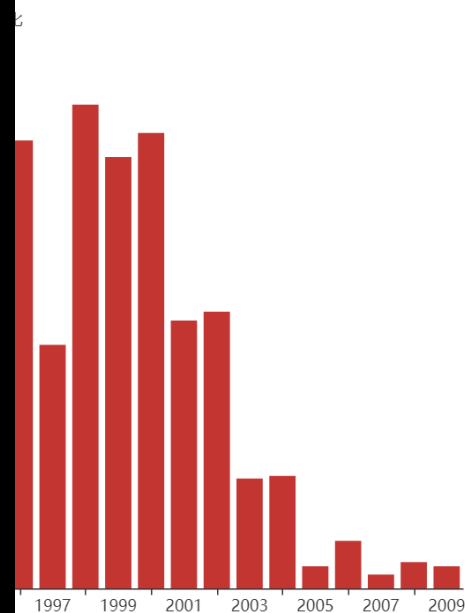
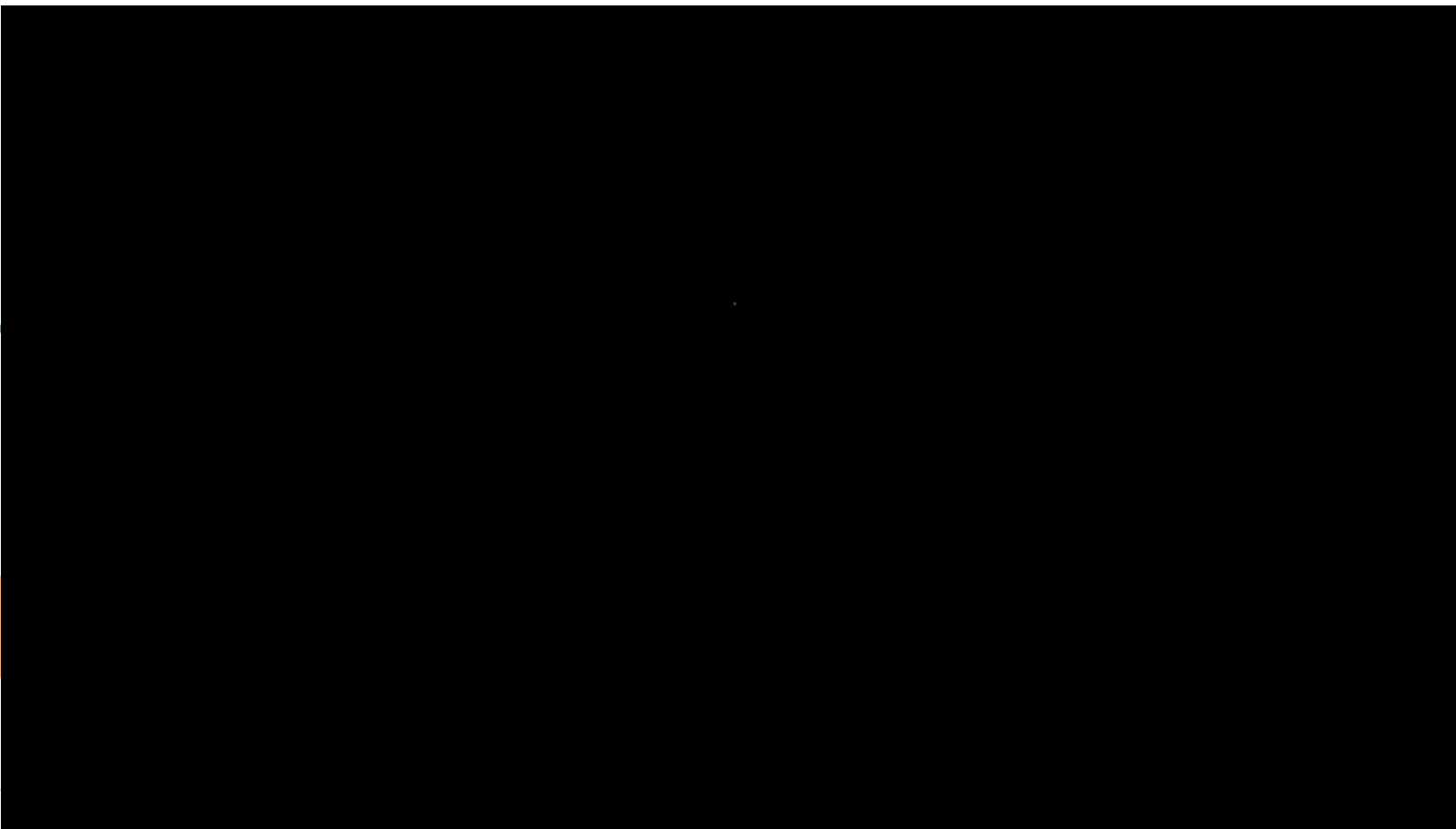
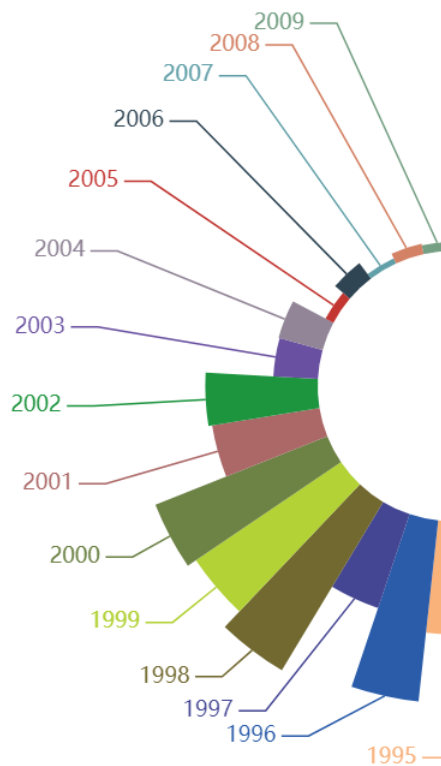


用户出生年份占比图

用户出生年份



数据分析及可视化

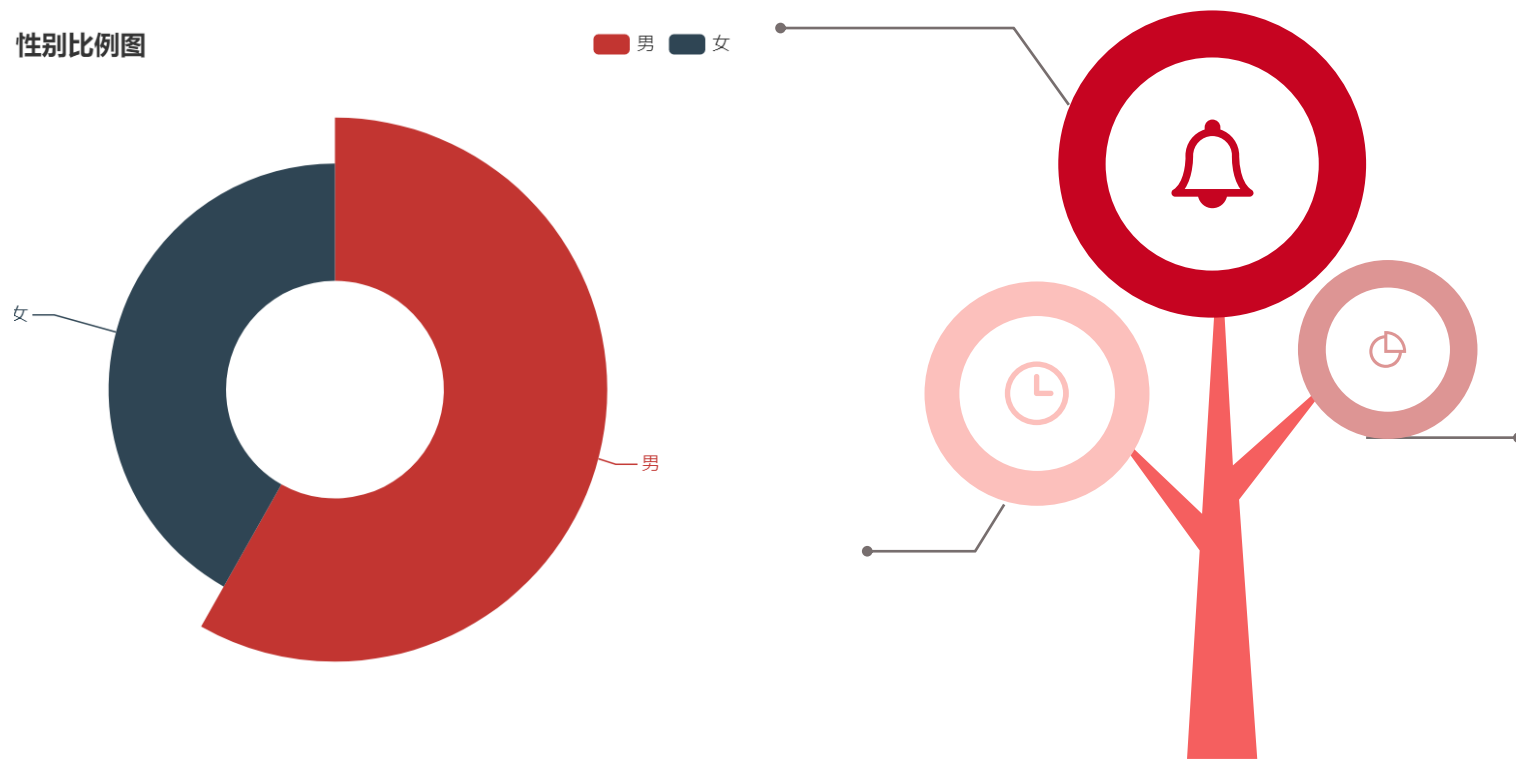


用户出生年份占比图

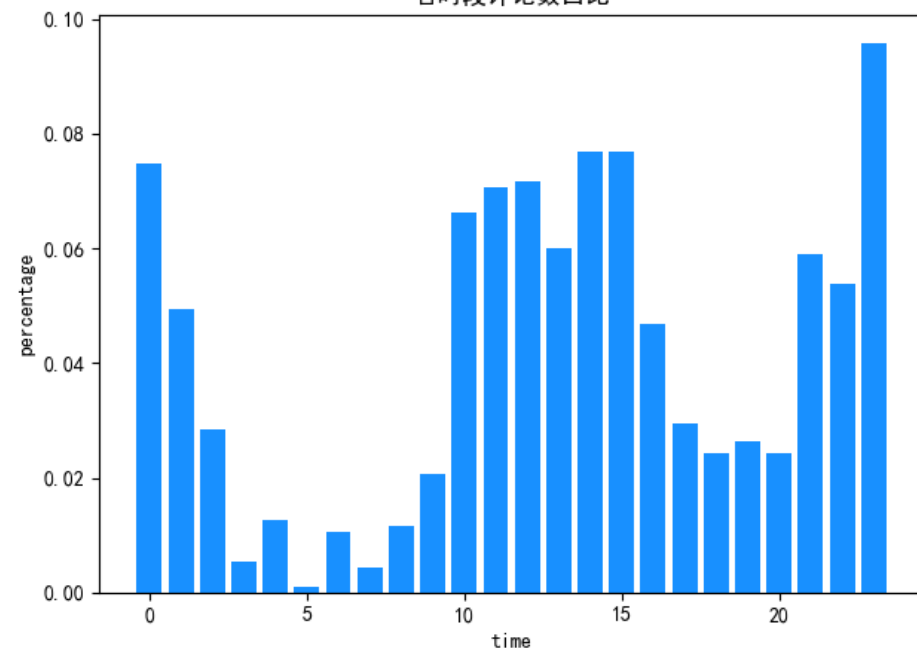


用户性别比例图

性别比例图



各时段评论数占比

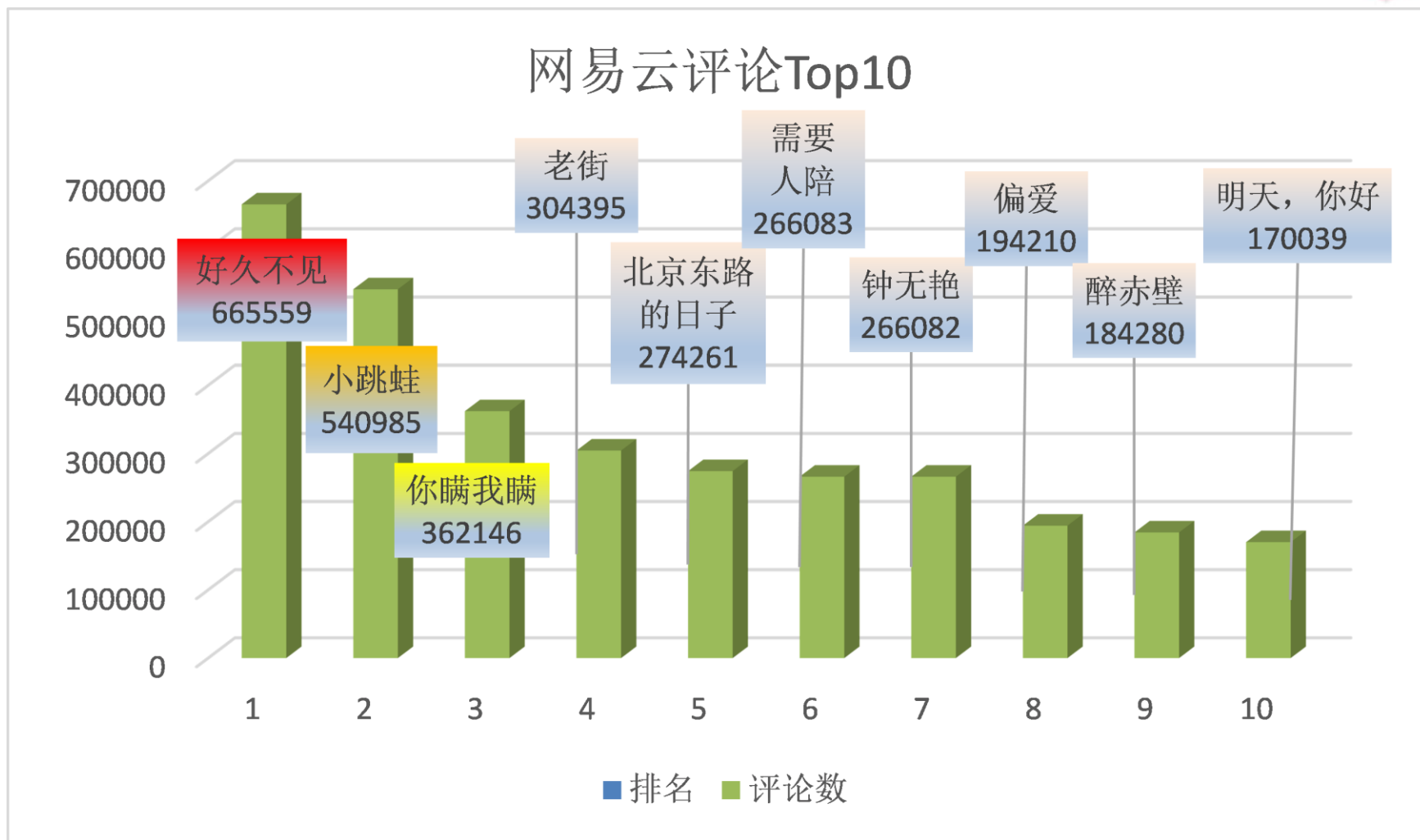


用户在各时间段评论占比

数据分析及可视化

网
易
云
音
乐
评
论

Top10



THANKS FOR YOUR LISTENING

感谢您的聆听

★ 汇报人：杜涵悦 ★

★ 学号：10181900144 ★