

小白UP主打怪 升级路线

HyperG1

李云帆 施俞晨 王玉世 邓淇升



目录



01 背景介绍

02 变量说明

03 描述分析

04 模型建立

05 结论建议



背景介绍





B站印象

产品定位

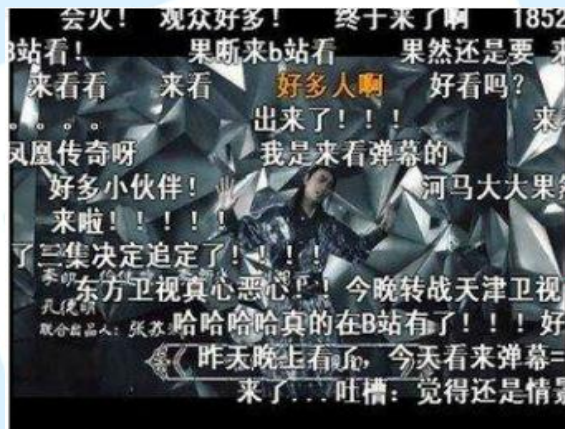
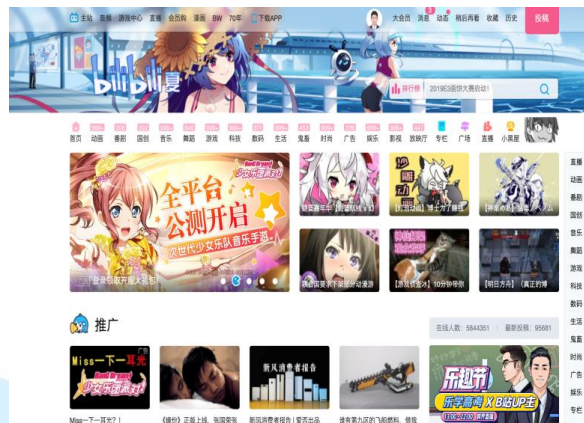
基于弹幕视频分享的互联网社区

弹幕功能

视频播放时画面上飘过的即时评论

ACG

Animation动画 + Comic漫画 + Game游戏





视频板块

10个板块：

动画、番剧、音乐、舞蹈、游戏、科技、娱乐、鬼畜、电影、电视剧

动画	MAD·AMV	MMD·3D	动画短片	综合			
番剧	连载动画	完结动画	资讯	官方衍生	国产动画	新番Index	
音乐	翻唱	VOCALOID·UTAU	演奏	三次元音乐	同人音乐	OP/ED/OST	音乐选集
舞蹈	ACG相关舞蹈	三次元舞蹈					
游戏	单机联机	网络游戏	电子竞技	音游	MUGEN	GMV	
科技	纪录片	趣味科普人文	野生技术协会	演讲·公开课	军事	数码	机械
娱乐	搞笑	生活	动物圈	美食圈	综艺	娱乐圈	Korea相关
鬼畜	二次元鬼畜	三次元鬼畜	人力VOCALOID	教程演示			
电影	欧美电影	日本电影	国产电影	其他国家	短片	电影相关	
电视剧	连载剧集	完结剧集	特摄·布袋戏	电视剧相关			



发展情况

月活跃用户量

5843万

- 相比2017年2月增长40.1%
- 相比2016年7月增长63.2%

日新增用户量

45.7万

- 每个周末出现高峰
- 节假日期间明显提升

注：数据来自“极光大数据”，收集截至2017年7月



深入挖掘

独特机制

B站采用“投稿 - 评分 - 排榜”机制进行数据管理

赞赏方式

观众可依个人喜好给出通过每日登陆获得的“硬币”或“积分”

推送方式

系统根据视频在一段时间内的评分高低给出日、周和月排行，并将最热门的视频推送至首页供所有用户观看

充电计划

在线打赏功能，鼓励用户为自己喜爱的UP主“充电”。



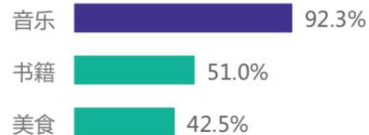
用户情况

哔哩哔哩app用户画像

年龄



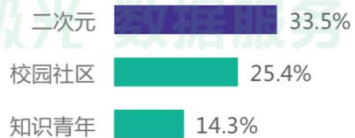
兴趣爱好



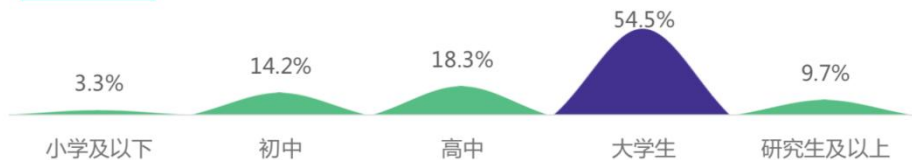
性别



社交风格



学历





如何做一个up主?





深入思考

有关B站发展的问題：

- 打造网红偶像？
- 新用户及运营？
- 不只是二次元？





变量说明





原始变量

变量类型	变量名	变量说明	取值范围
因变量	播放量	离散变量	整数：817 – 37128221，单位：次
自变量 (视频信息)	视频标题	文本数据	例：A路人做声线测试，最后疯了
	视频时长	离散变量	整数：0 – 17129，单位：秒
	视频号	分类变量（28459水平）	每个视频的视频号独一无二，例：8741476
	频道	分类变量（60水平）	例：单机游戏、美食圈、手机平板等
	标签	文本数据	例：#攻略##黑桐谷歌##PS4#
	点赞	离散变量	整数：0 – 1528233，单位：次
	投币	离散变量	整数：0 – 1944322，单位：个
	收藏	离散变量	整数：0 – 1423719，单位：次
	分享	离散变量	整数：0 – 873886，单位：次



原始变量

变量类型	变量名	变量说明	取值范围
自变量 (up主信息)	up主	分类变量 (101水平)	例: 老番茄、敖厂长、LexBurner等
	性别	分类变量 (3水平)	男、女、保密
	粉丝数	离散变量	浮点数: 67.3 – 528.1, 单位: 万人
	类型	分类变量 (16水平)	例: 鬼畜、动画、游戏等
	充电	离散变量	整数: 268 – 91191, 单位: 次
	是否有专栏	分类变量 (2水平)	0、1
	投稿时间	时间变量	例: 2019/2/1 13:30
	周几	分类变量 (7水平)	1、2、3、4、5、6、7

- 经过筛选分析，主要挑选加粗变量进行分析与建模，其他变量作为附属数据予以显示



描述分析





播放量

播放量

衡量视频受欢迎程度的一个重要指标

播放量	结果
最小值	817
最大值	3700万

跨度太大

对数变换

对数播放量	结果
最小值	6.7
最大值	17.4

➤ 播放量最小的视频:

粉丝数	up主	视频名
269.5万	起小点是大腿	L星人snake经理左雾: 就算0.1%的帮助我也会去做

➤ 播放量最大的视频:

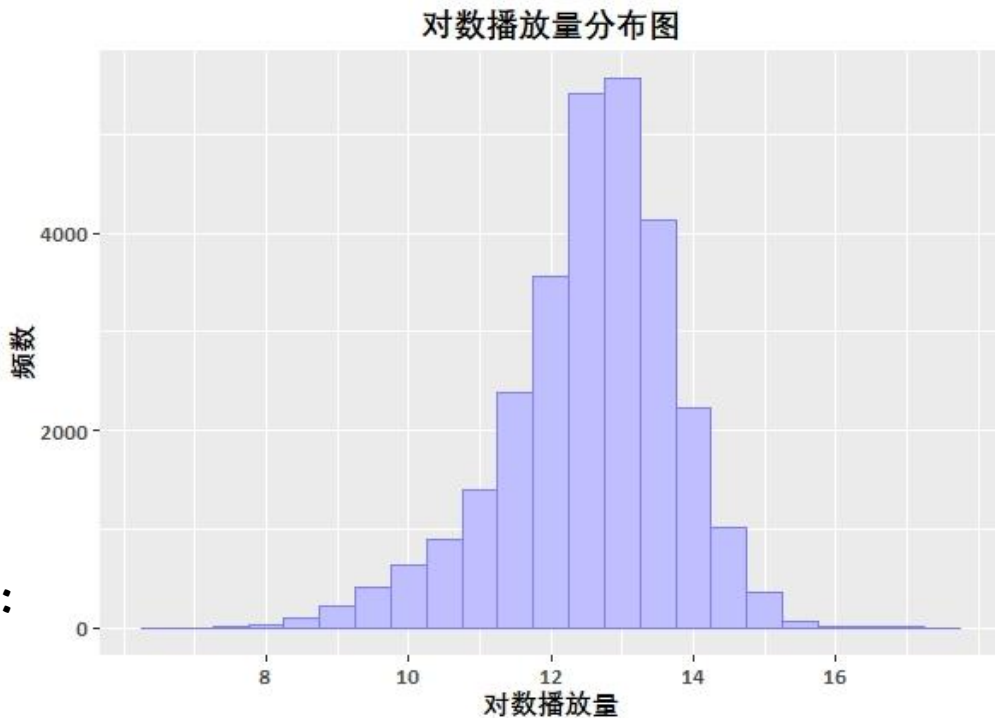
粉丝数	up主	视频名
125万	小可儿	【春晚鬼畜】赵本山: 我就是念诗之王! 【改革春风吹满地】



对数播放量

对数播放量	结果
最小值	6.7
最大值	17.4
均值	12.5
中位数	12.7

- 直方图呈右偏趋势
- KS检验显著拒绝了正态分布的原假设
- 大部分视频的播放区间：
对数播放量：12-14
播放量：16万-120万





用户“三连”

用户“三连”

点赞 + 投币 + 收藏



6.1万



不喜欢



2.8万



3.7万



1.5万

- 代表了视频被用户的认可程度
- 对up主的人气和知名度十分重要。

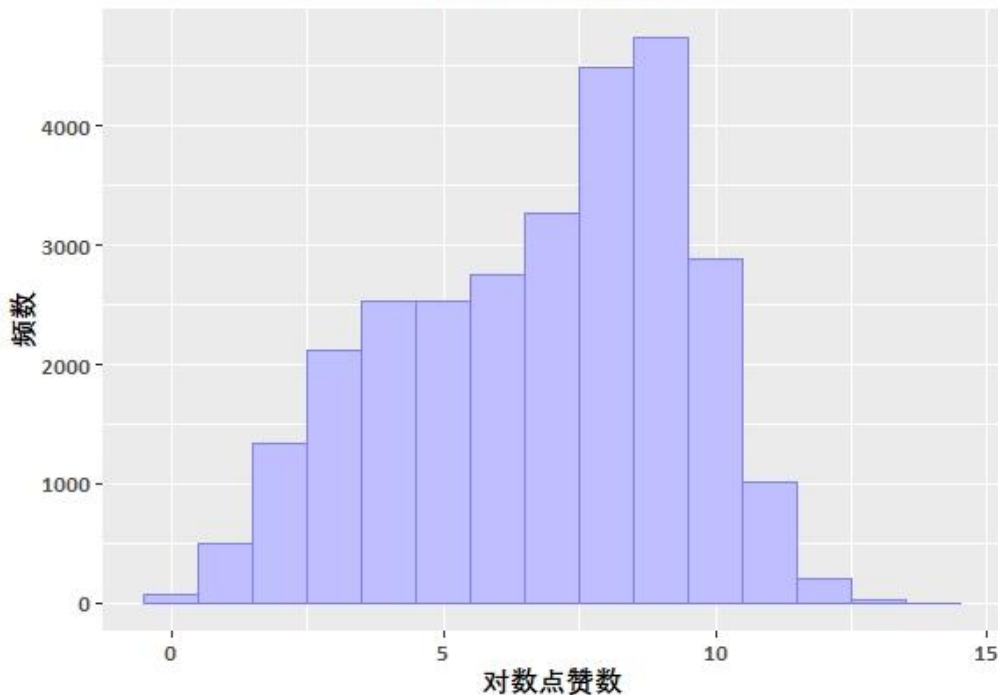


点赞数

对数点赞数	结果
最小值	0
最大值	14.2
均值	6.8
中位数	7.3

- 有67个视频点赞数为0
- 有关“赵本山”的视频点赞数最多

对数点赞数分布图



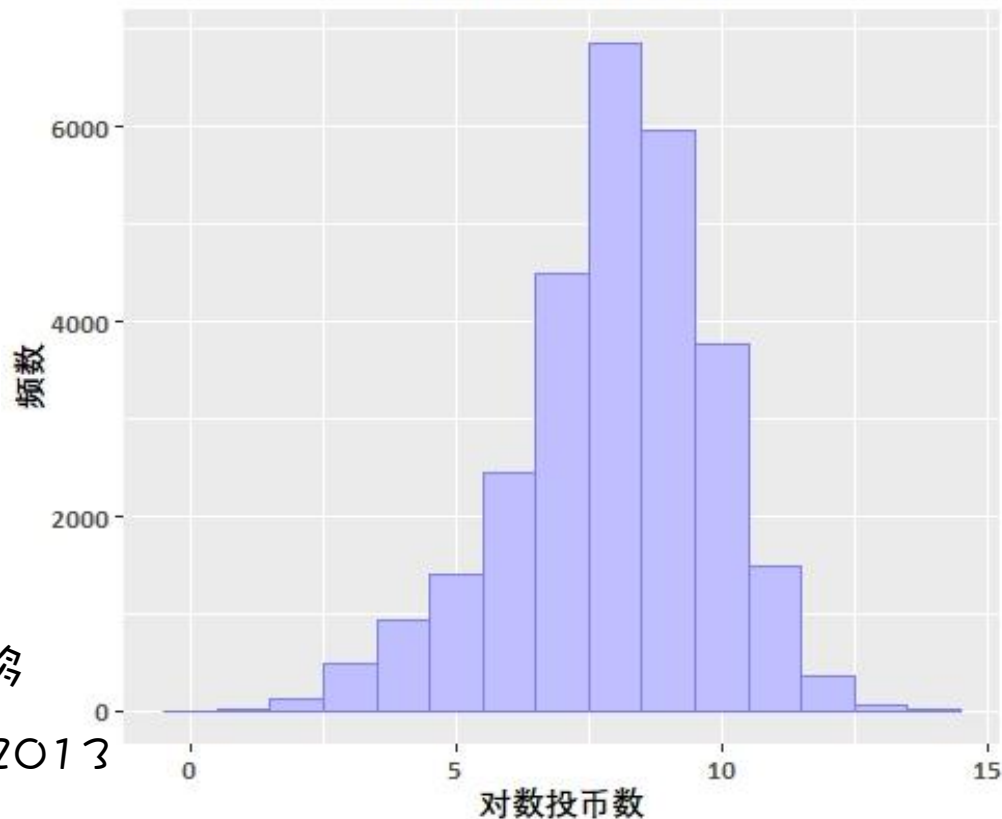


投币数

对数投币数	结果
最小值	0
最大值	14.5
均值	8.0
中位数	8.1

- 直方图呈右偏趋势
- KS检验显著拒绝了正态分布的原假设
- 投币数为0的视频为“超级多玩星期天14期花絮”，于2013年发布，播放量仅有6424次

对数投币数分布图



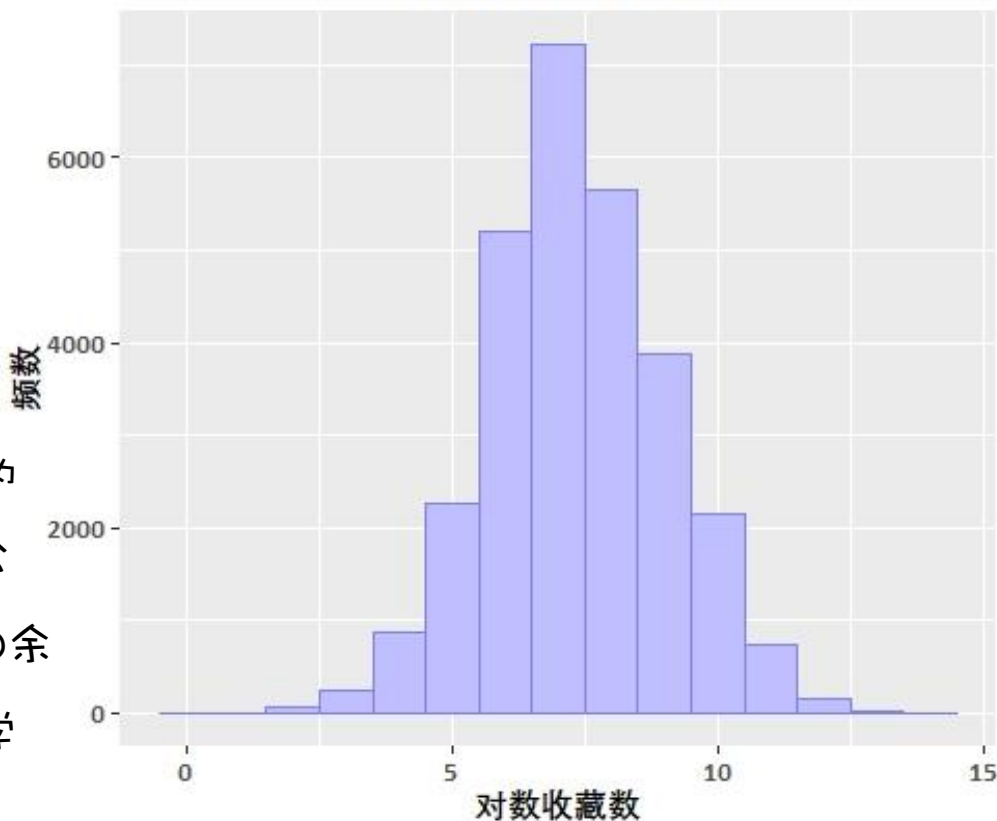


收藏数

对数收藏数	结果
最小值	0
最大值	14.2
均值	7.4
中位数	7.3

- 收藏数为0的视频为up主“长歌与小见见”发布的“【LOL】我是人鱼小公举”，播放量仅有3000余次，该视频可能缺乏教学意义，因此无人收藏

对数收藏数分布图



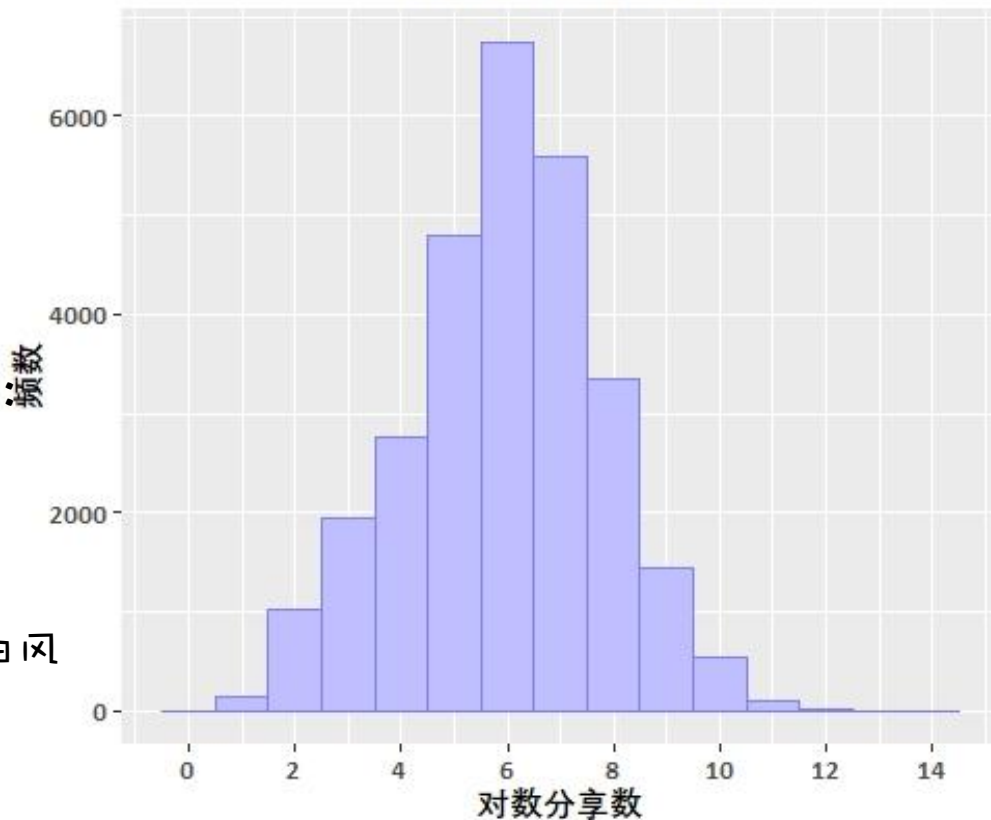


分享数

对数收藏数	结果
最小值	0
最大值	13.7
均值	5.9
中位数	6.0

- 大部分视频的分享数区间：
对数分享量：5-8
分享量：148-2981
- 分享数为0的7个视频由抽风crazy、神奇陆夫人、中国boy等几位up主发布。

对数分享数分布图





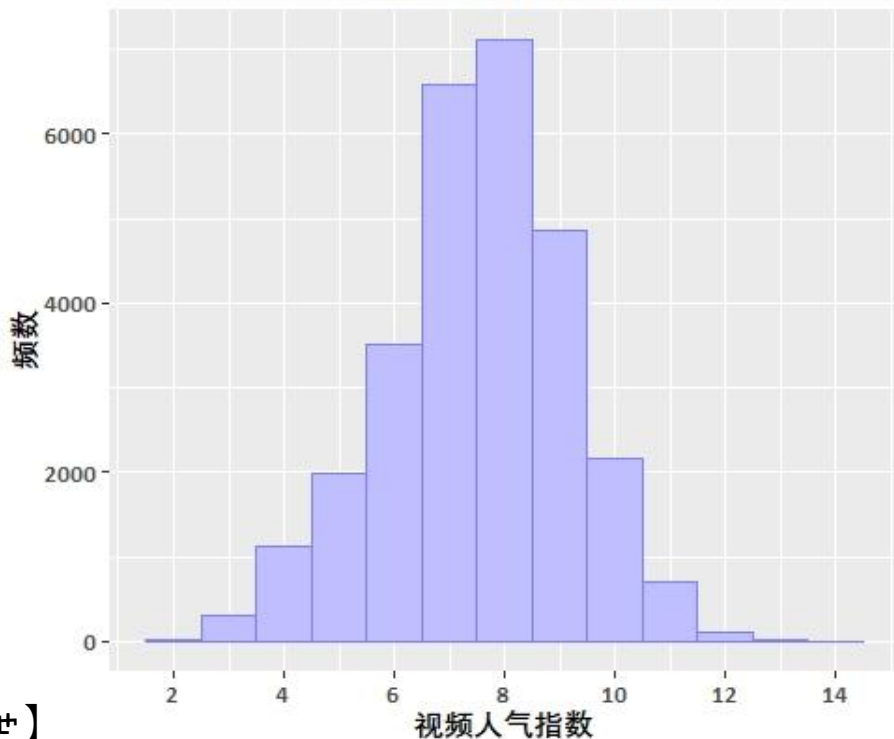
视频人气指数

视频人气指数 = 收藏 $\times 0.3$ + 分享 $\times 0.3$ + 投币 $\times 0.2$ + 点赞 $\times 0.1$ + 播放 $\times 0.1$

视频人气指数	结果
最小值	1.7
最大值	14.4
均值	7.5
中位数	7.6

- 视频人气指数最低：
L星人snake经理左雾：就算
0.1%的帮助我也会去做
- 视频人气指数最高：
【春晚鬼畜】赵本山：我就
是念诗之王！【改革春风吹满地】

视频人气指数分布图

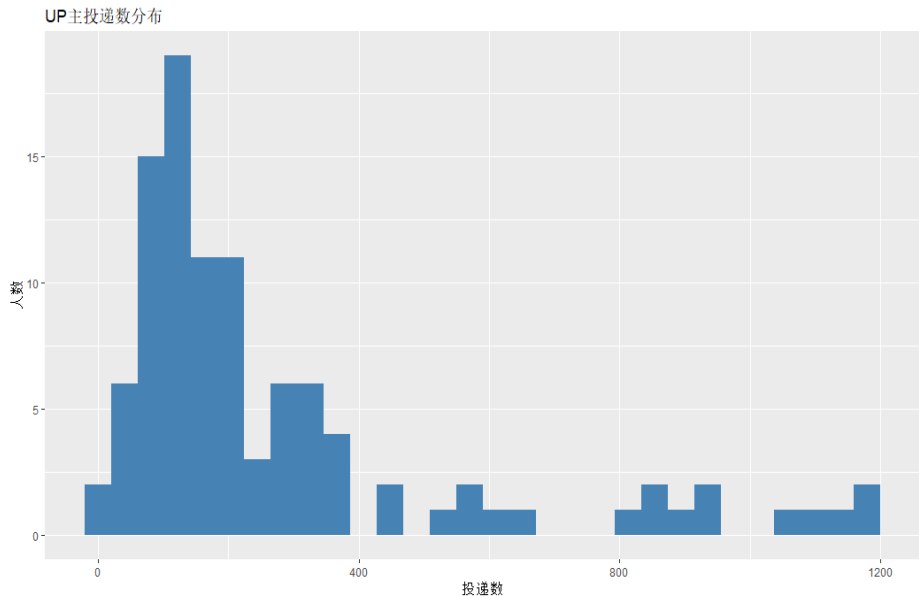




up主人气指数

up主人气指数 = 视频人气指数 \times 30 + 粉丝数 \times 20 + 充电数 \times 25

up主	up主人气指数
敖厂长	10466
LexBurner	9696
老番茄	7840
中国BOY 超级大猩猩	7608
起小点是大腿	7564
木鱼水心	6421





老番茄

关注数

1

粉丝数

461.3万

播放数

3.8亿



- 复旦金融系大四学生
- 著名B站UP主
- 知名网络播客和游戏娱乐解说人
- 粉丝数：367万
- 自制视频总播放量：2.9亿
- 成名作：《你可见过如此丧心病狂的口袋妖怪解说》系列。

“你可以在老番茄身上看到自己的影子。老番茄就是我们的同龄人，你可以看见他的成长和精进，不管是学业上的、生活上的还是视频上的。他代表你实现了人生的一部分可能性。”



搜索指数 ?

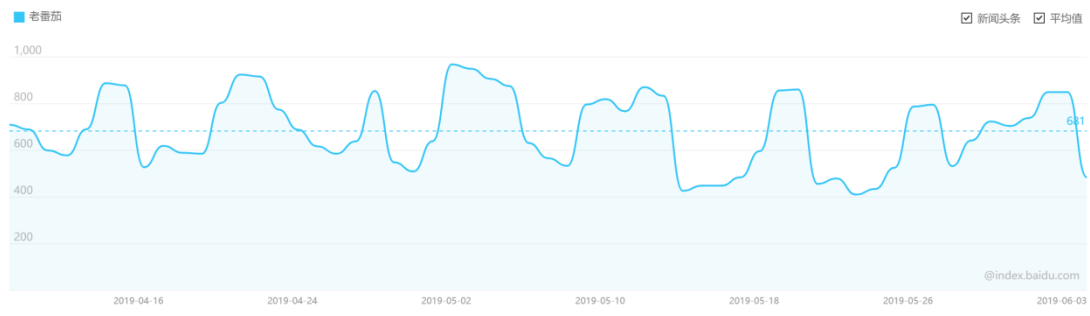
2019-04-08 ~ 2019-06-03 | 自定义 | PC | 全国 |



PC端搜索指数

搜索指数 ?

2019-04-08 ~ 2019-06-03 | 自定义 | 移动 | 全国 |



移动端搜索指数



分区王者

目标

分析视频人气指数较高的各频道视频

频道

- 共有60个视频频道
- 超过500个视频的频道有13个
- 汇总上述每个频道视频人气指数最高的5个视频，共得到65个热门视频

频道	视频数
单机游戏	6877
电子竞技	4410
搞笑	2451
美食圈	1871
影视杂谈	1690
日常	1576
综合	1431
翻唱	1292
网络游戏	1108
鬼畜调教	988
美妆	855
手机游戏	630
手机平板	508



分区王者

up主	热门视频数	备注
爱做饭的芋头SAMA	6	5个美食圈频道，1个日常频道
神奇的老皮	6	4个电子竞技频道，2个搞笑频道
LexBurner	4	4个均为综合频道，但内容为动画点评
科技美学	4	专注于手机平板评测
老番茄	4	专注于单机游戏解说
机智的党妹	2	无
某幻君	2	无
努力的Lorre	2	无
渗透之C君	2	无
伊丽莎白鼠	2	无



分区王者

通过此分析，也希望
初入B站的萌新
可以根据上述榜单
找到自己心仪的up

视频分区	分区王者
单机游戏	老番茄
电子竞技	神奇的老皮
翻唱	A路人
搞笑	神奇的老皮、徐大虾
鬼畜调教	伊丽莎白鼠
美食圈	爱做饭的芋头SAMA
美妆	机智的党妹
日常	信誓蛋蛋
手机平板	科技美学
手机游戏	敖厂长、泛式、老番茄、芒果冰OL和渗透之C君
网络游戏	某幻君
影视杂谈	努力的Lorre
综合区	LexBurner



热门视频

【老番茄】史上最骚杀手(第一集)

【真人炉石】暗影崛起-最秀新卡发布！

为什么老外把中餐送到宇宙里？

papi酱突然更新的放送——《我妈爱say no！》全网火爆上线~

手搓冰粉到底有什么不一样？这个夏天冰粉一定要拥有姓名！！

英文版《改革春风吹满地》！！Chinese people so 牛逼！

【一花】超啰嗦的新手妆教程 / 重点是痘痘肌的底妆分享

【中文八级】两个国人展开了惊人的英语八级对话

剧情刺激！场面爆笑！这谁顶得住啊！2019一月新番吐槽大盘点！

【春晚鬼畜】赵本山：我就是念诗之王！【改革春风吹满地】

【游戏侦查冰】揭秘游戏氪金抽卡的“黑幕”

「科技美学直播」拼多多爱爱“旗舰手机体验，那岩已经疯了”

《复联4》前必看！一口气看完21部漫威电影，完整的时间线剧情讲解！



模型建立





模型说明

1. 线性回归

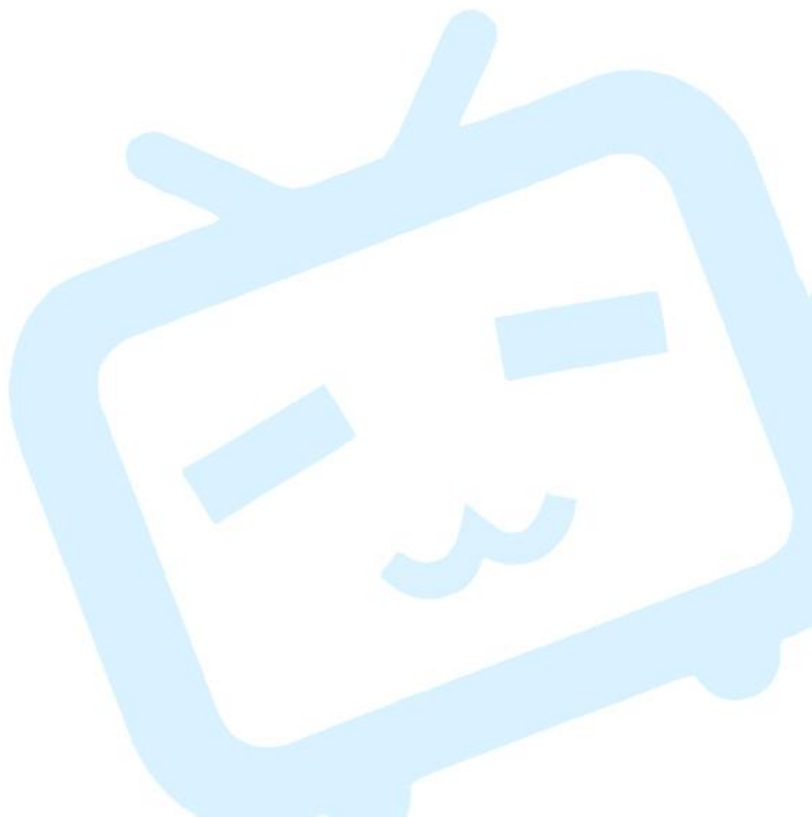
2. LASSO

3. 回归树

4. XGBoost

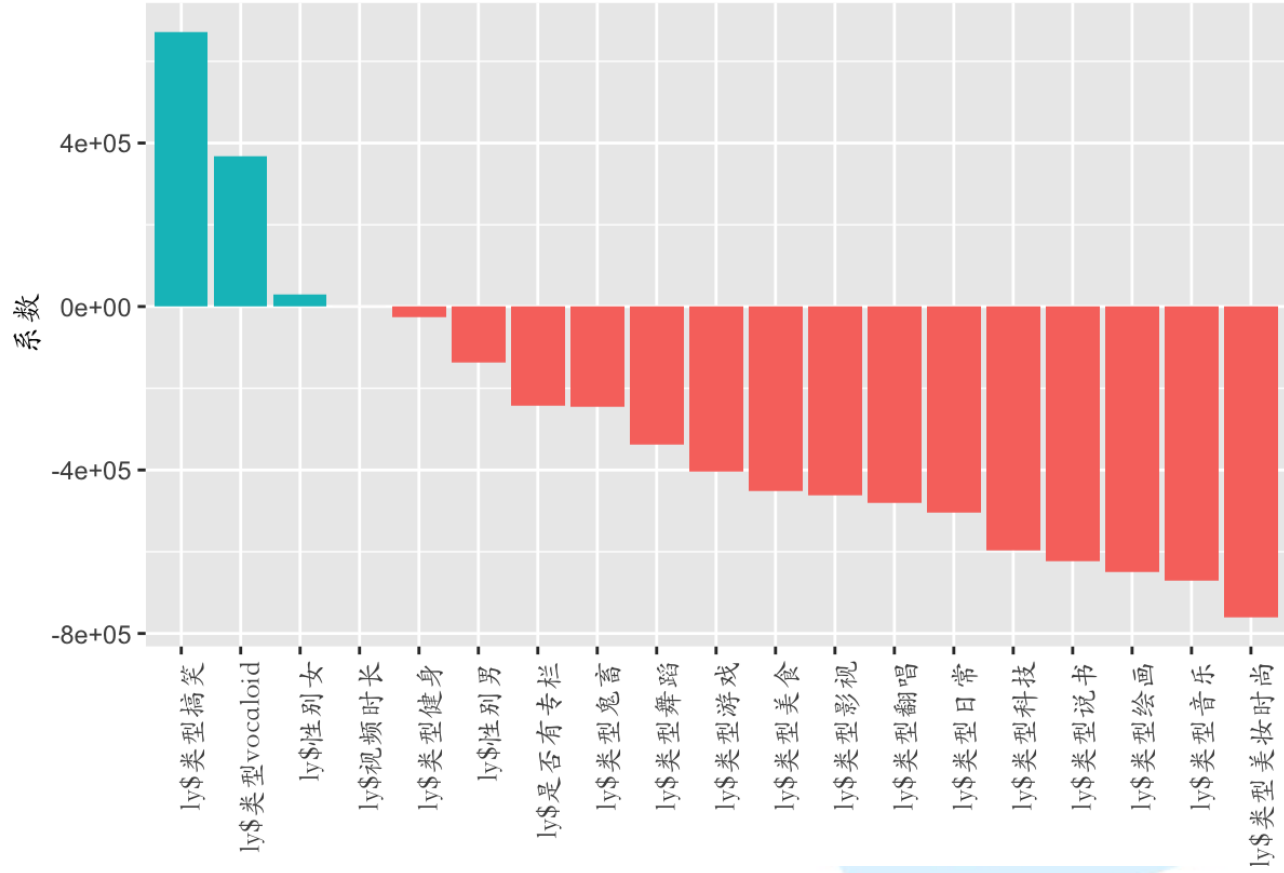
5. 聚类分析

6. 视频标题研究



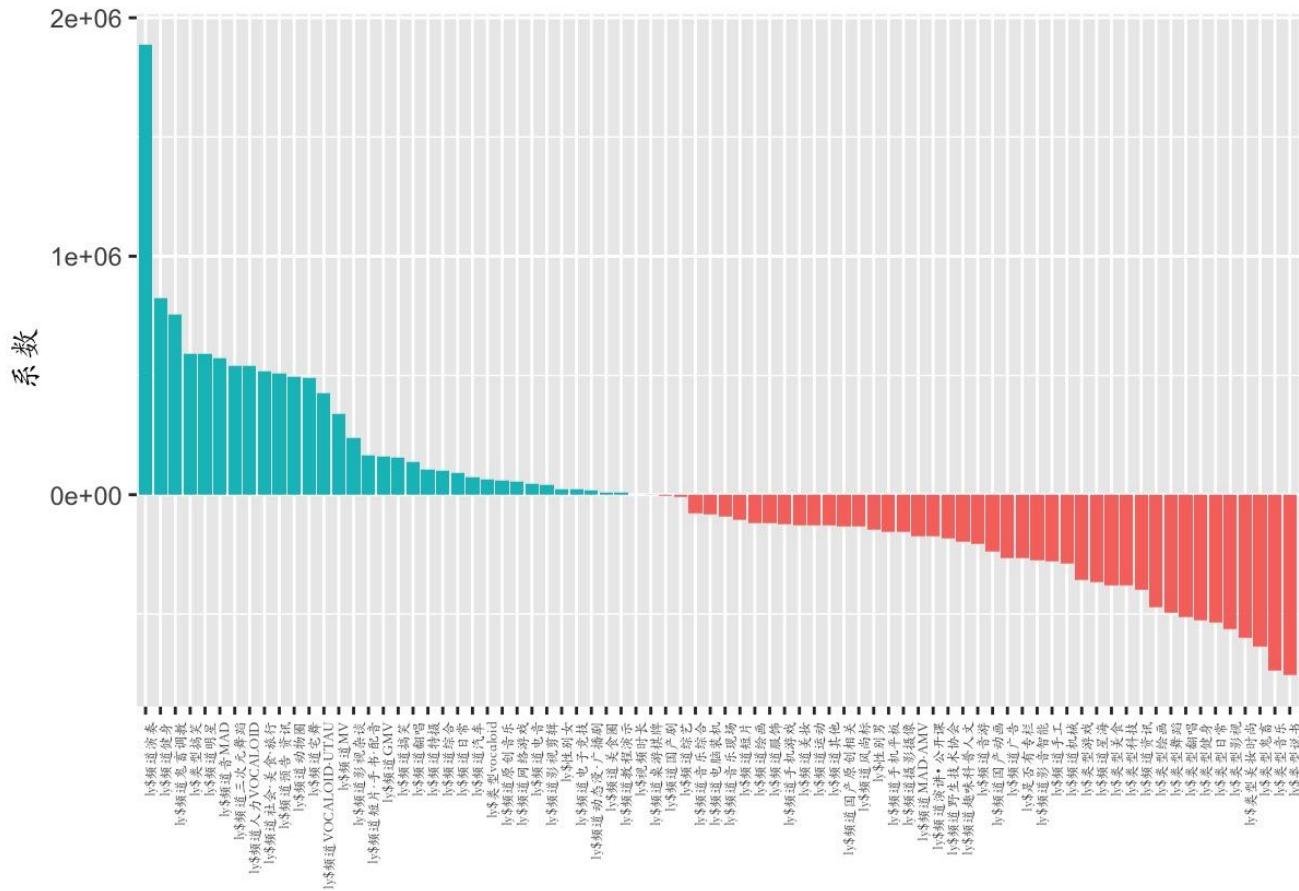


线性回归





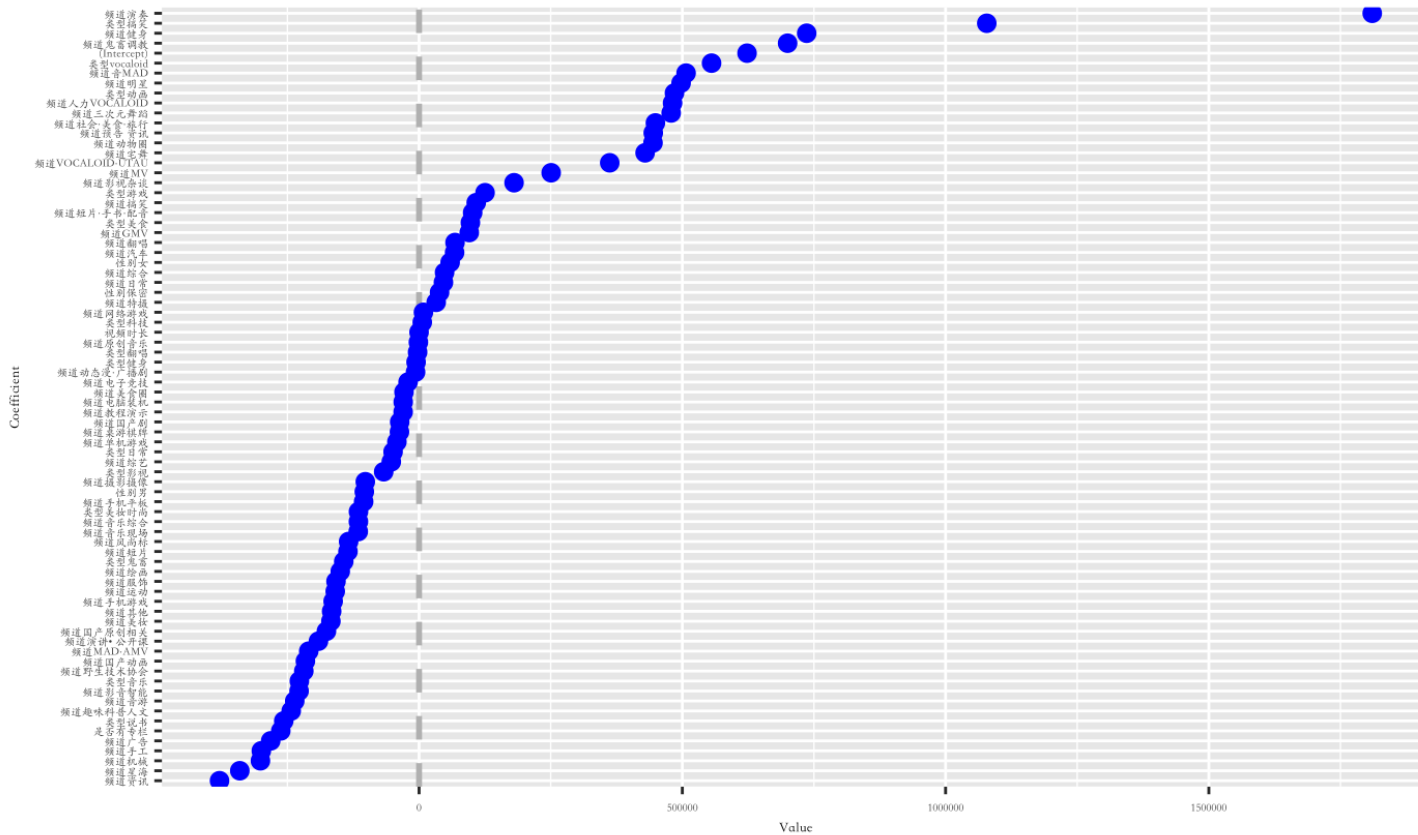
线性回归





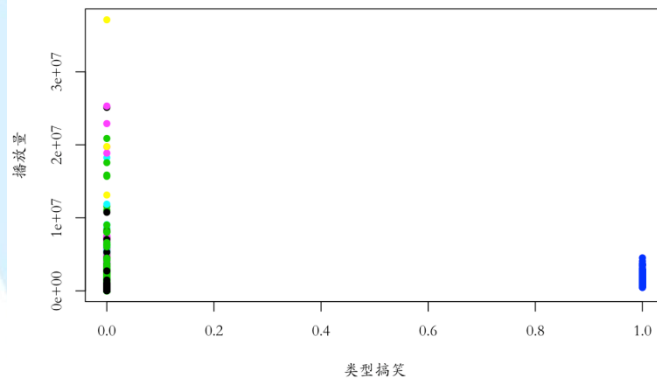
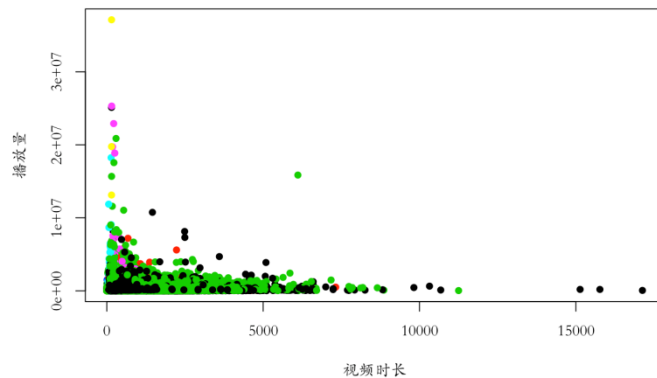
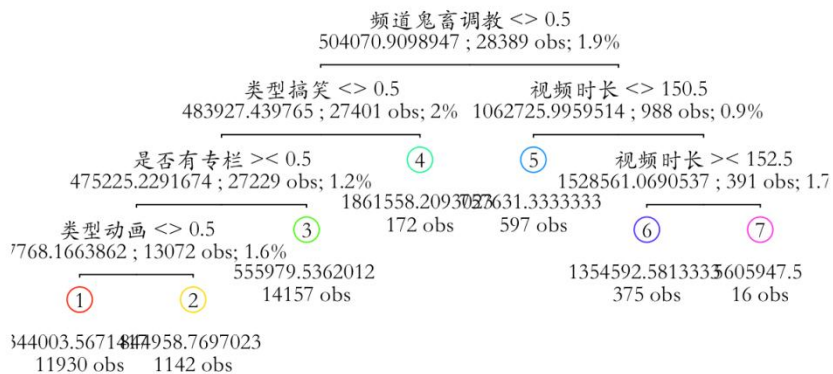
LASSO

Coefficient Plot





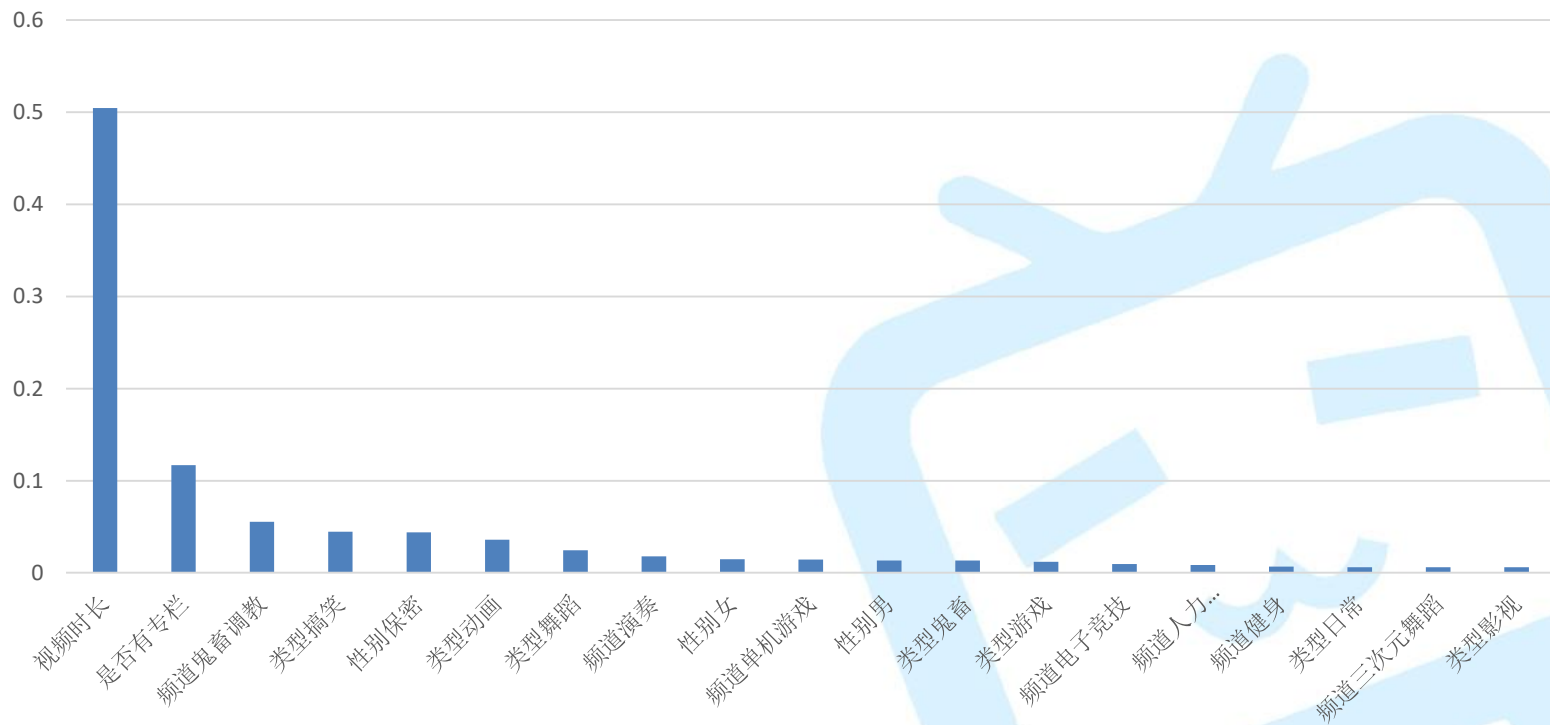
回归树





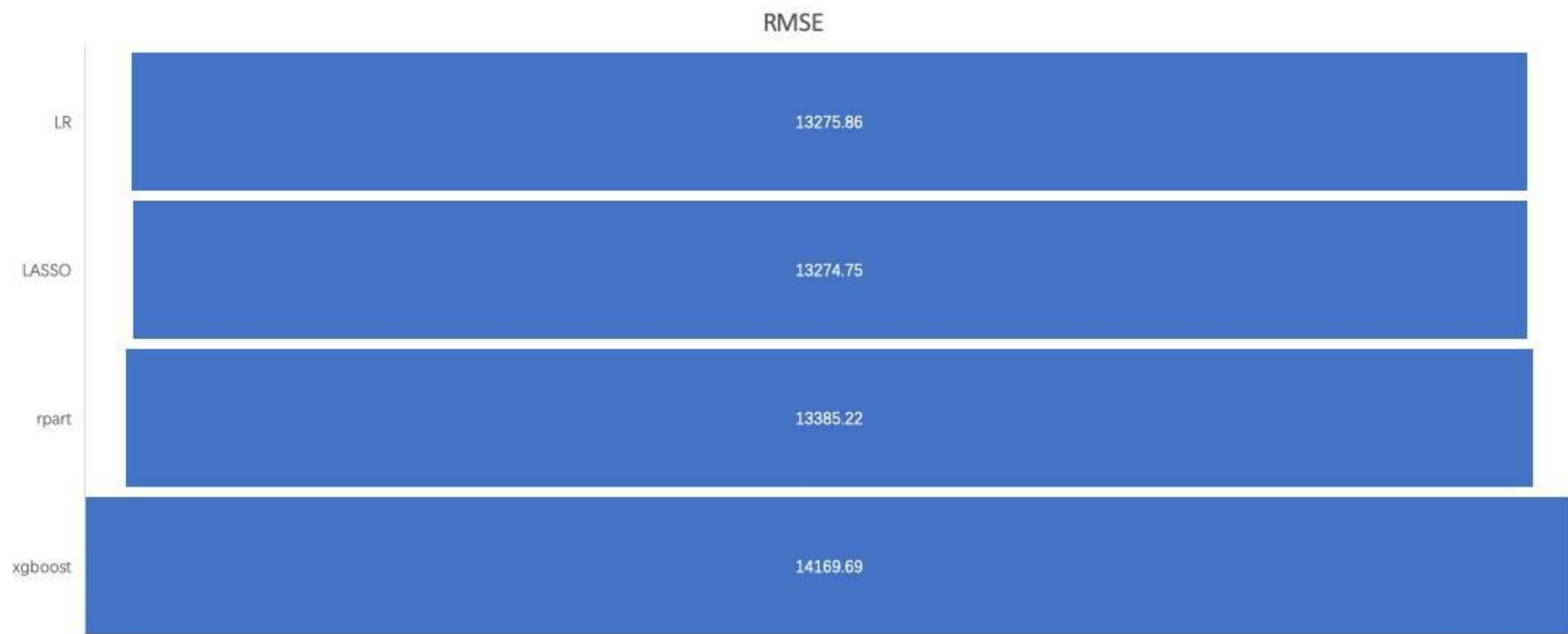
XGBoost

Gain For XGBoost



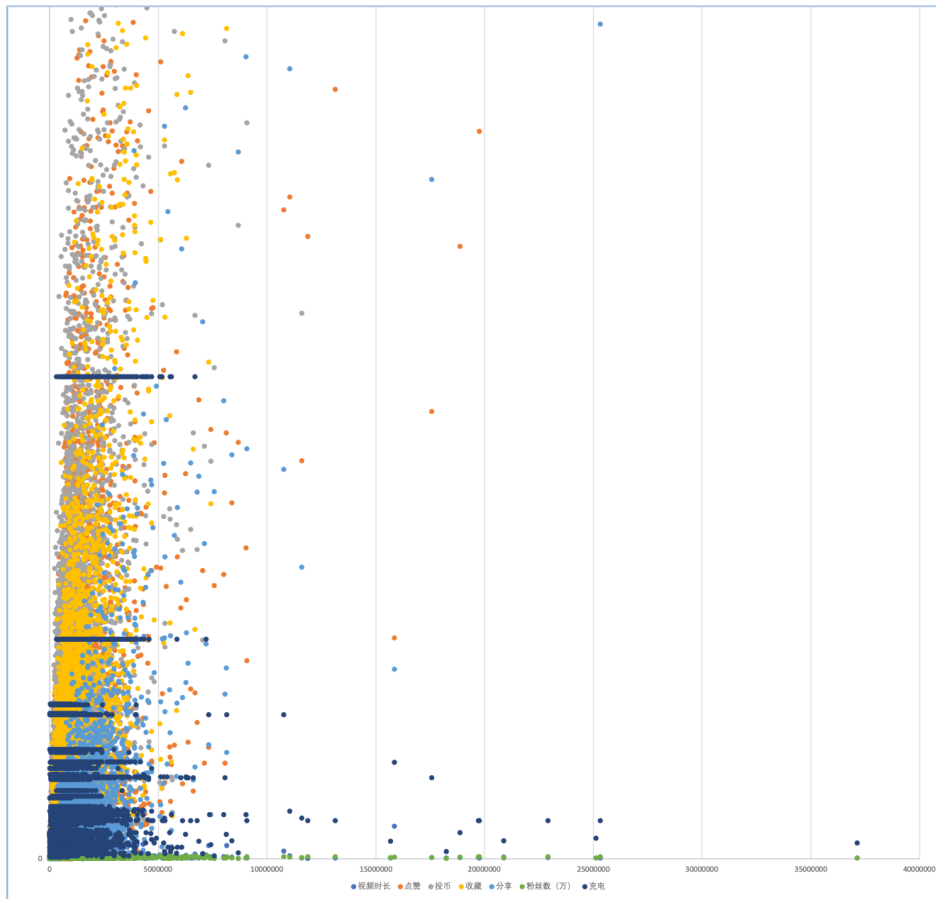
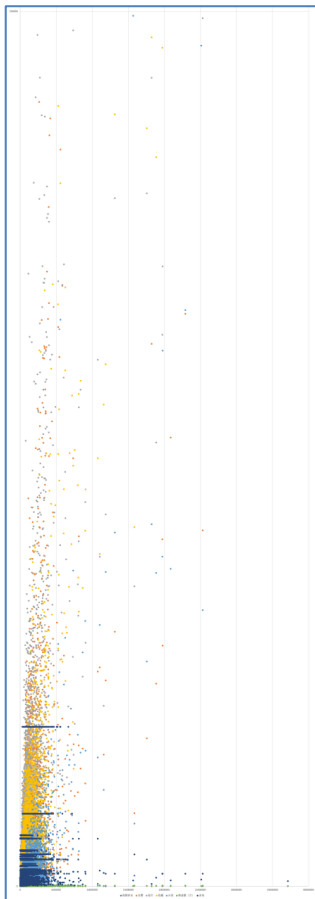


回归结果





关于播放量的散点图





聚类分析

观察播放量的散点图可知：

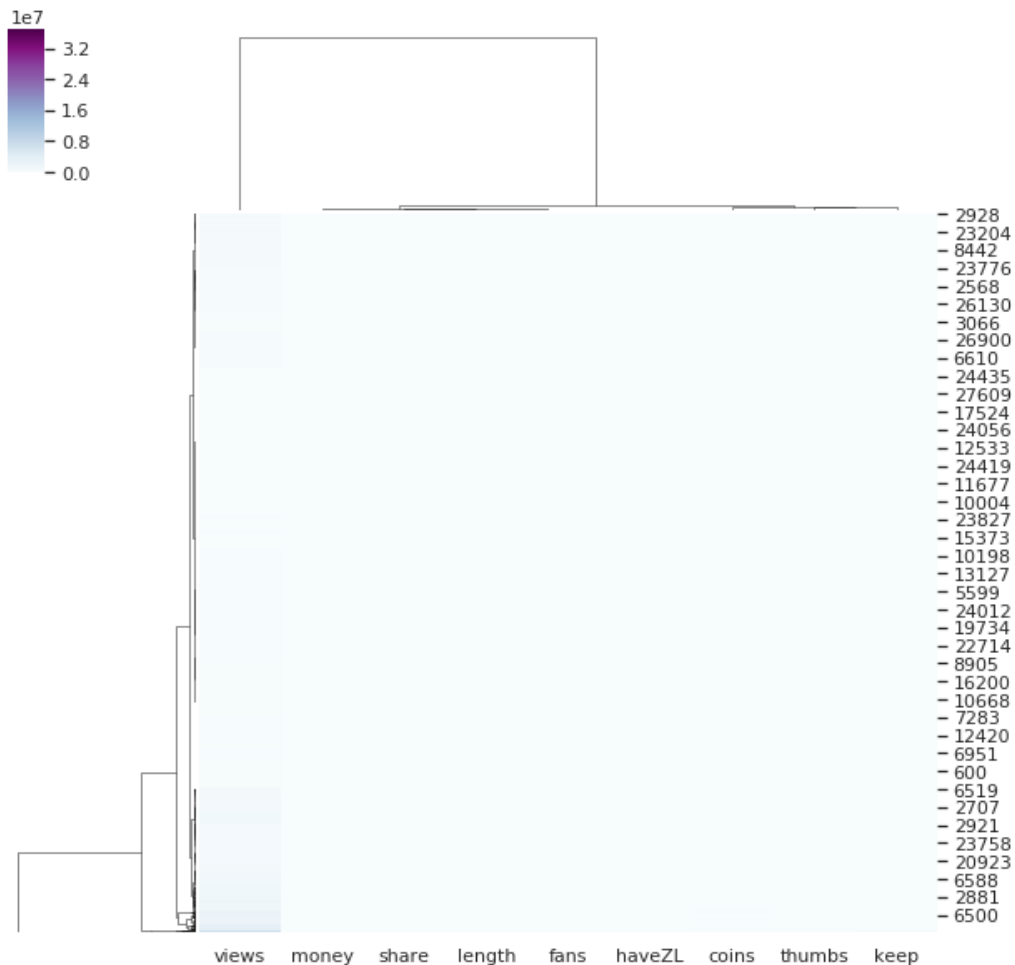
1. 各个变量在播放量较低处较为集中，但又有播放量与点赞，投币，分享这几个变量有正相关。
2. 粉丝数在我们的观察中比较稳定在百位数量级。
3. 视频时长也相对来说波动较小。

散点图展示了各个变量很有特色的分布，我们可以对各个变量进行聚类分析！



聚类分析

数据未进行归一化
时的聚类结果

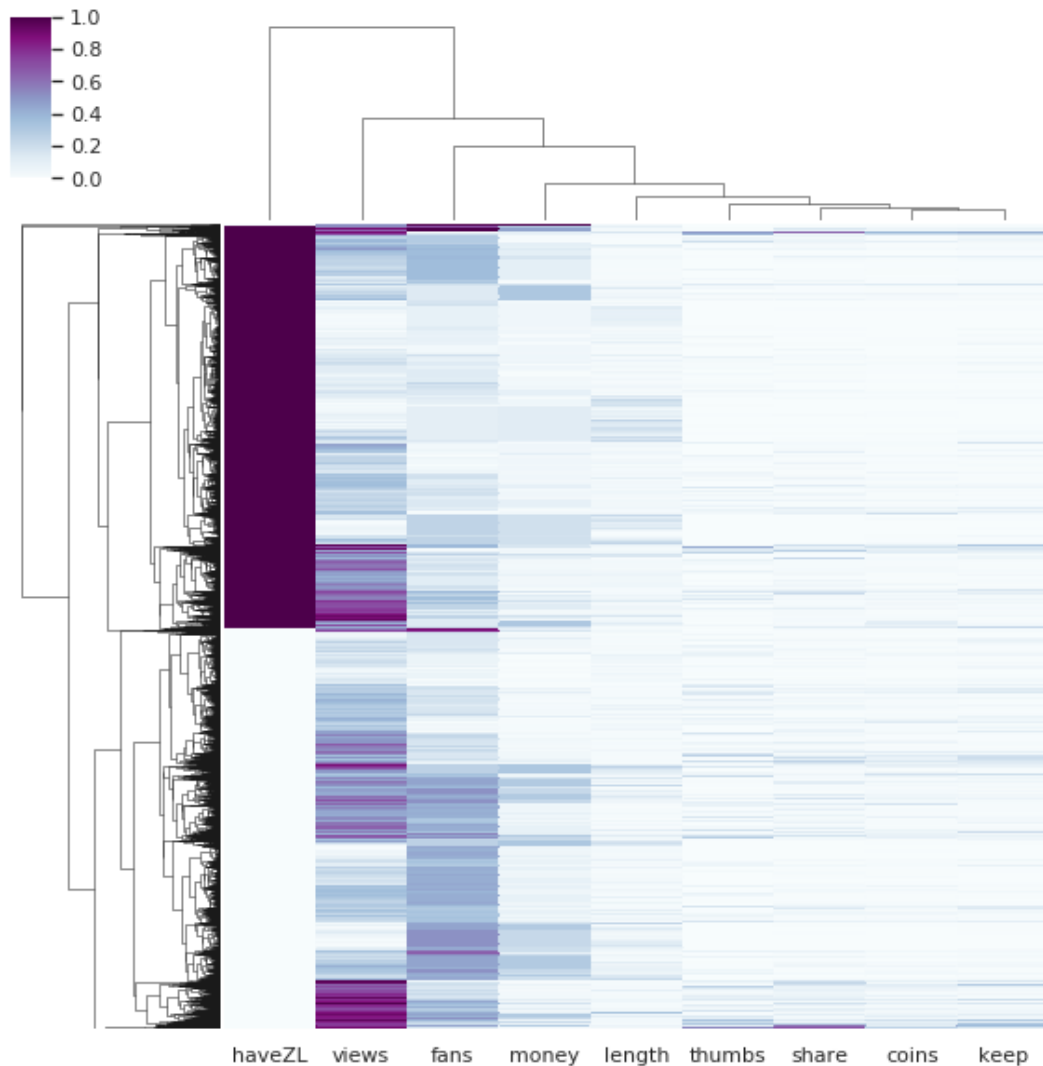




聚类分析

取部分数值类型变量
来研究聚类效果：

变量	描述
haveZL	是否有专栏
Views	播放量
Fans	粉丝数
Money	充电
Length	视频时长
Thumbs	点赞
Share	分享
Coins	投币
Keep	收藏





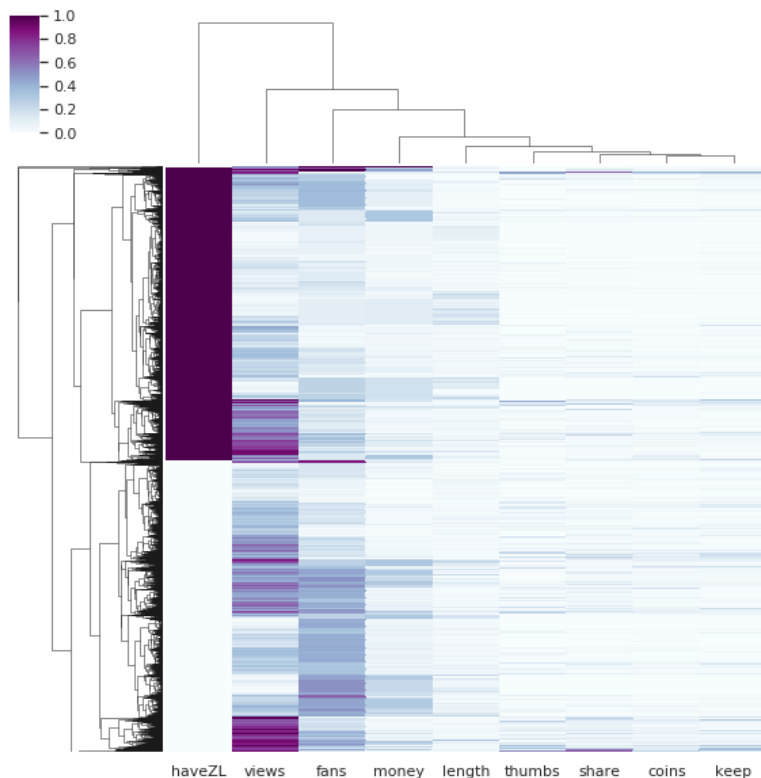
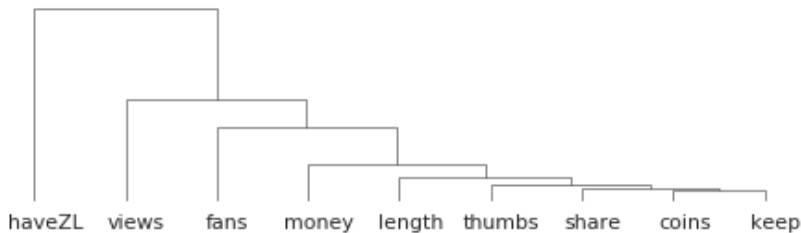
聚类分析

数据进行归一化后的聚类结果

1. 几个变量按照顺序依次形成
聚类（极度不平衡？）

2. 各个视频的分散程度高，但
在宏观上又有某些共同特性

接下来进一步分析视频的标题





视频标题的文本分析

吃	0.024
笑点	0.019
咬人	0.015
鸡	0.014
毛蛋	0.013
♥	0.013
	0.012
舞蹈	0.009
做	0.007
爱	0.006
西四	0.006
♂	0.006
ω	0.006
Benny	0.006
分享	0.006
欣小萌	0.005
全明星	0.005
番茄	0.004
黑暗	0.004
第一次	0.004

可爱吃货区

杂谈	0.049
主播	0.031
一分钟	0.029
篇	0.026
TOP10	0.023
玩	0.022
小点	0.021
真会	0.014
绘画	0.008
老师	0.008
死	0.007
粉丝	0.007
野食	0.006
青铜	0.006
修炼	0.006
手册	0.006
徐	0.005
超神篇	0.004
鬼	0.004
小哥	0.004

标题党

	0.049
直播	0.012
游戏	0.011
「	0.01
天天	0.009
开箱	0.009
谷歌	0.009
黑桐	0.008
科技	0.008
美学	0.007
喵	0.006
中国	0.006
手绘	0.006
试玩	0.005
猫	0.005
新	0.005
体验	0.004
君	0.004
美	0.004
解说	0.004

天天看直播

游戏	0.038
守望	0.016
先锋	0.014
VS	0.01
恐怖	0.009
小强	0.009
文曰	0.009
分钟	0.009
作	0.008
陆夫人	0.008
读	0.007
故事	0.006
wanna	0.006
火影忍者	0.006
神奇	0.006
原著	0.006
	0.005
说	0.005
猫	0.005
王老菊教	0.004

游戏对战



视频标题的文本分析

期	0.029
木鱼	0.017
剧场	0.014
盘点	0.014
翻唱	0.013
炸	0.012
微	0.011
小缘	0.011
站	0.009
B	0.008
龙珠	0.008
C	0.007
炉石	0.007
英雄	0.006
素材库	0.006
周六	0.006
局长	0.006
萧忆情	0.005
撒	0.005
痒	0.005

小众

六道	0.101
绅士	0.028
中国	0.015
散人	0.013
解说	0.013
时刻	0.011
视频	0.011
BOY	0.009
日本	0.007
新番	0.007
小熊	0.007
集锦	0.007
方言版	0.007
猫	0.006
四川	0.006
推荐	0.006
动画	0.006
游戏	0.006
实况	0.006
挑战	0.006

文化

童年	0.027
经典	0.02
秀	0.013
欢乐	0.011
鬼畜	0.01
期	0.01
世界	0.008
抽风	0.008
努巴尼	0.007
游戏王	0.006
LOL	0.006
水彩	0.005
长歌	0.005
厨房	0.005
画	0.005
版	0.005
开心	0.005
躺	0.004
十大	0.004
污妖王	0.004

2013青年欢乐多

Latent Dirichlet Allocation模型

该模型在英文文本上效果不错，不过在我们的数据集上貌似还有改进的空间。

可以看到这个话题模型的聚类结果还是比较纷繁的。



结论建议





结论建议

对于新进入B站大家庭的小白up主，如何才能马上变网红，走上人生巅峰？

1. 首先，选择自己的up主定位很重要，选择做一个游戏up或者鬼畜up可能能让你很快火起来哦～
2. 其次，视频的时间要控制一下哦，虽然长视频可能更有内涵，但别忘了受众都是年轻朋友鸭！做做精彩的短视频吧，你在B站的未来一片光明啊！
3. 如果看到自己精心制作的的视频播放量上不去，而一些知名up随便糊弄的视频都播放量爆表也不要过于沮丧啦，这可能是因为他们们的粉丝基数大，因而被推荐的几率也会更高啦～

THANKS FOR
LISTENING !



HyperG1