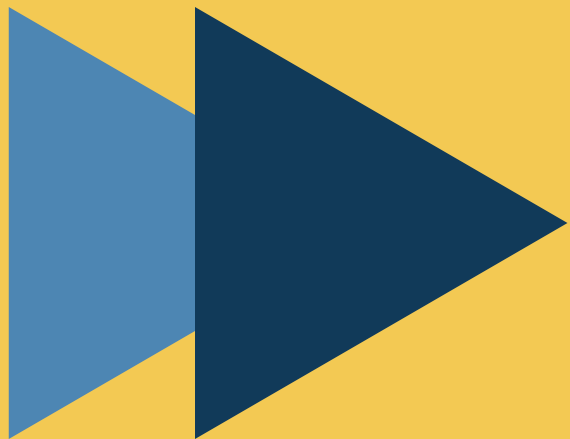


TALK
SHOW

脱口秀演员是怎么火起来的？

——商务大数据期末报告（第八组）

康蓝月 章程程 聂大庆 刘建华 胡帅帅



CONTENTS



1.背景介绍



2.数据介绍



3.模型构建



4.结论分析

1

背景介绍

1.1 脱口秀是什么？

脱口秀（Talk Show，中文名字是英文音译），也称为谈话节目，是指一种由观众聚集在一起讨论主持人提出的话题的广播或电视节目，是目前比较流行的一种喜剧类型。



1.2 脱口秀在中国的发展

谈话节目为
主的初始阶
段

突出个人品
牌的成长阶
段

以群体为中
心的创新阶
段

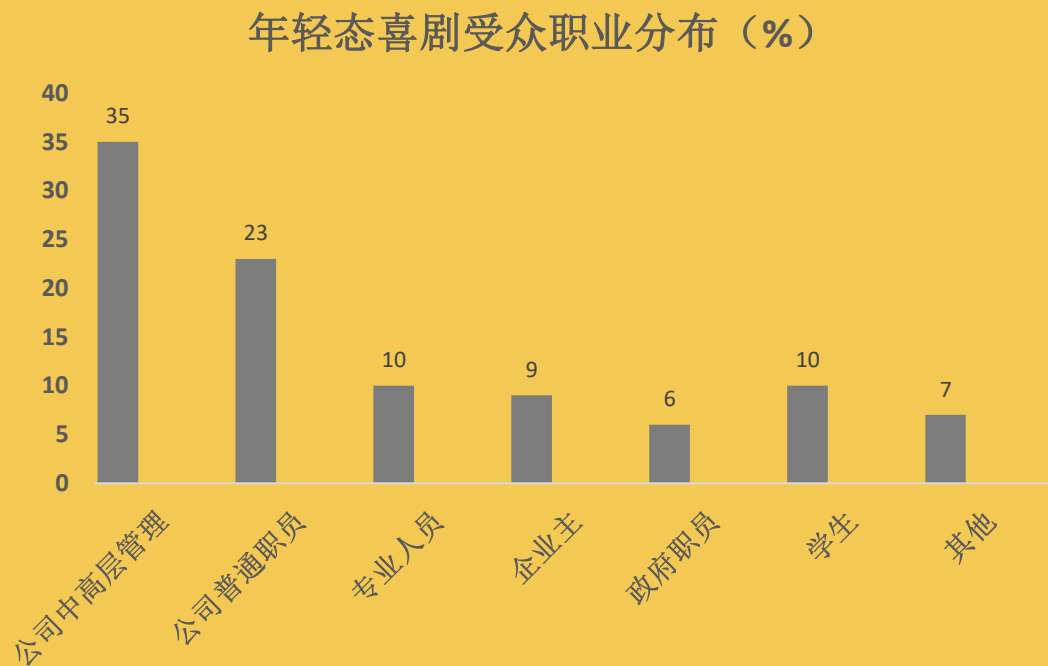
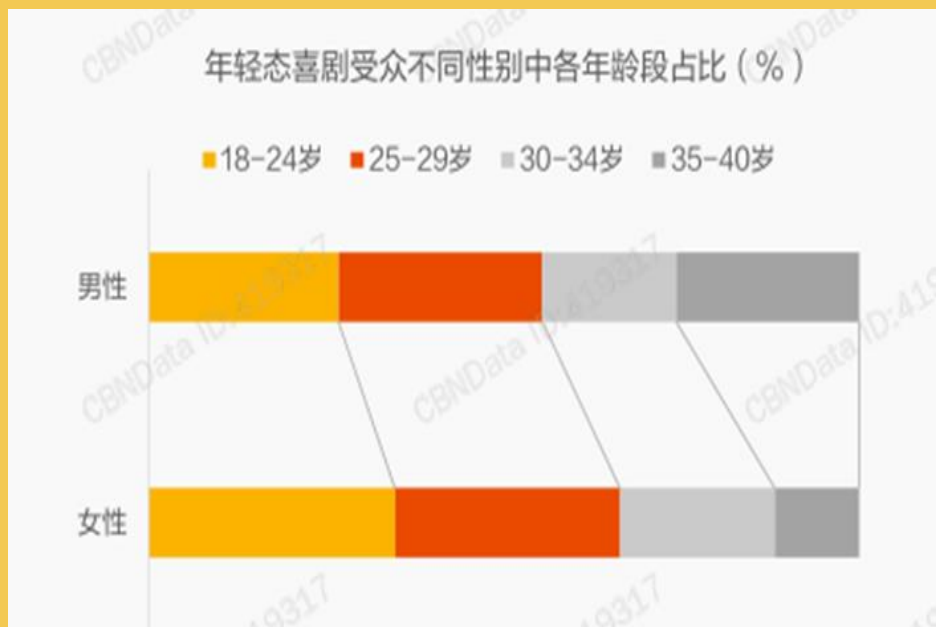
线下线上相互
补充的流量阶
段

阶段	时间	代表节目
第一阶段	1993年—2001年	1993年《东方直播室》，1996年《实话实说》 2000年《对话》，2000年《超级访问》， 2001年《鲁豫有约》
第二阶段	2002年—2013年	2005年《今晚》，2007年《大鹏嘚吧嘚》 2012年《今晚80后脱口秀》 2012年《小崔说立波秀》，2013年《暴走大事件》
第三阶段	2014年—2019年	2014年《奇葩说》，2015年《金星秀》 2017年《吐槽大会》，2017年《脱口秀大会》
第四阶段	2020年至今	2020年《脱口秀大会》第三季，《脱口秀反跨年》 2021年《脱口秀大会》第四季



1.3 脱口秀受众分析

年龄分布与职业分布：

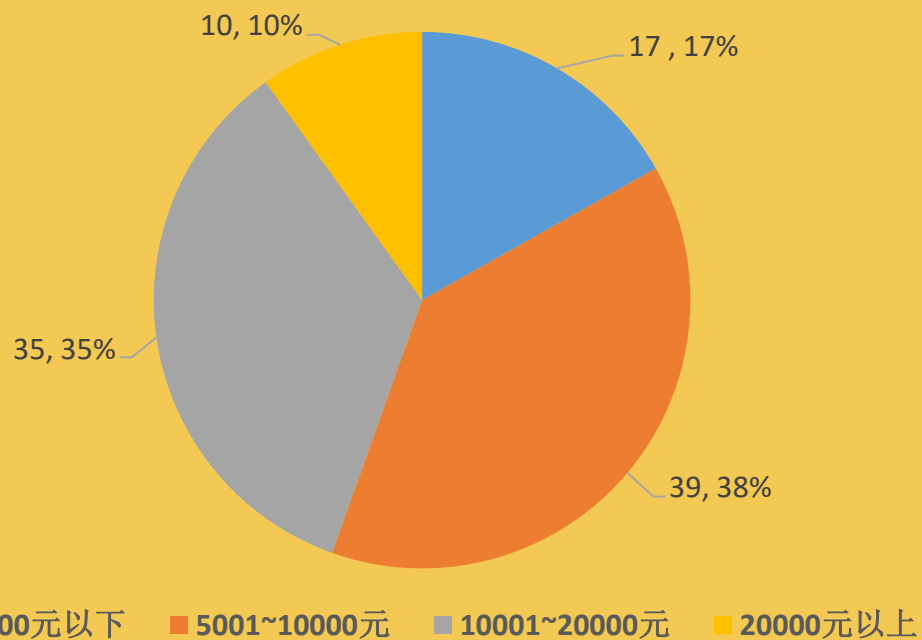


年轻态喜剧受众以18-29岁为主体，女性受众中年轻人偏多
职业主要集中在公司中高层和普通职员

1.3 脱口秀受众分析

收入状况：

年轻态喜剧受众个人月收入情况（%）



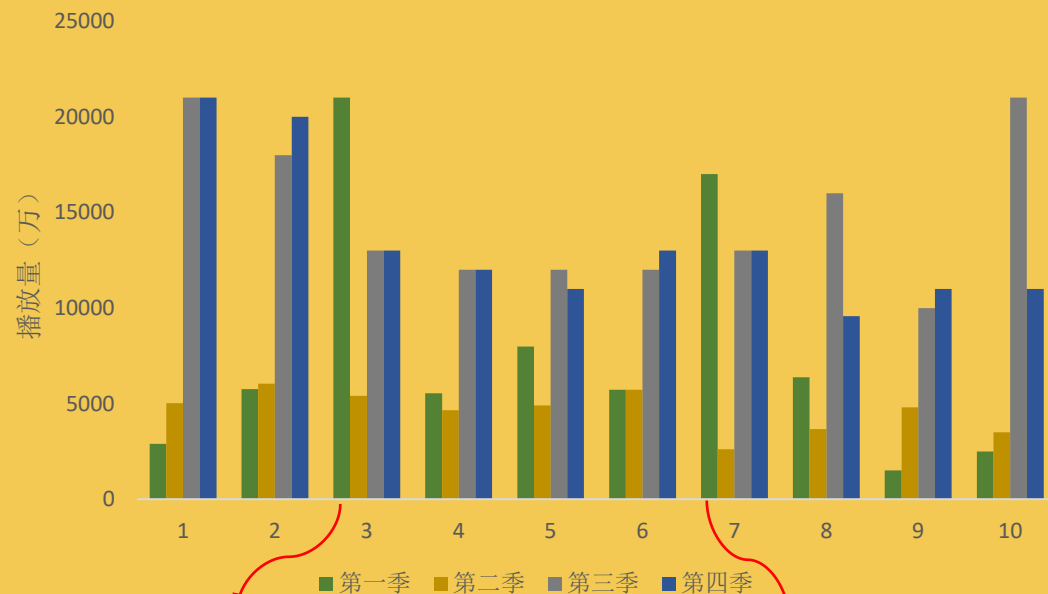
大部分受众的收入可观，购买力强



1.4 《脱口秀大会》节目分析

脱口秀大会四季播放量与搜索数：

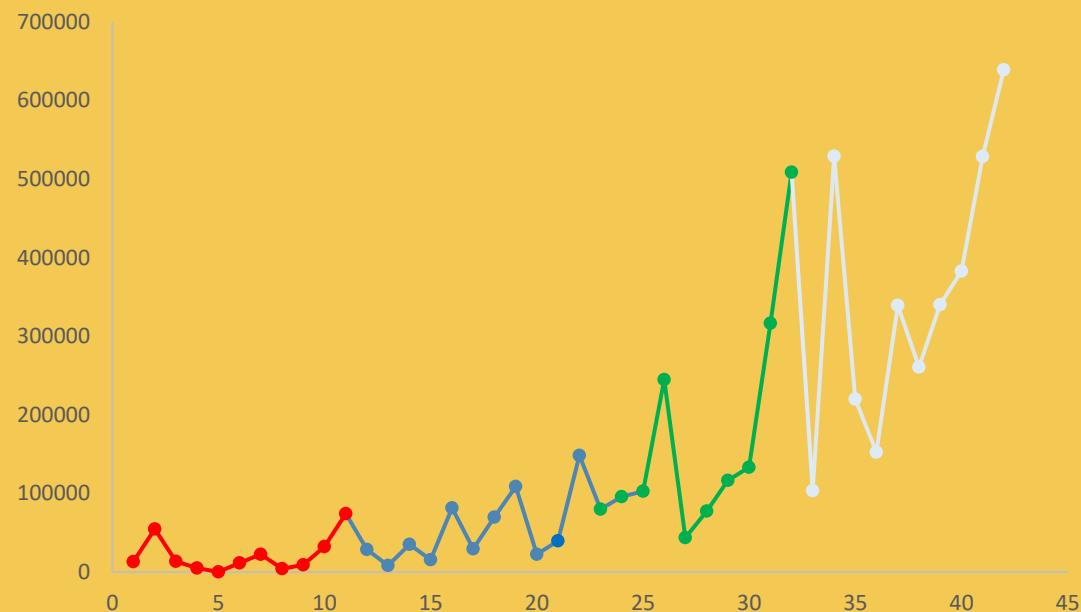
四季播放量对比图



沈梦辰和杜海涛恋情爆料

俞灏明烧伤后复出综艺秀

搜索指数

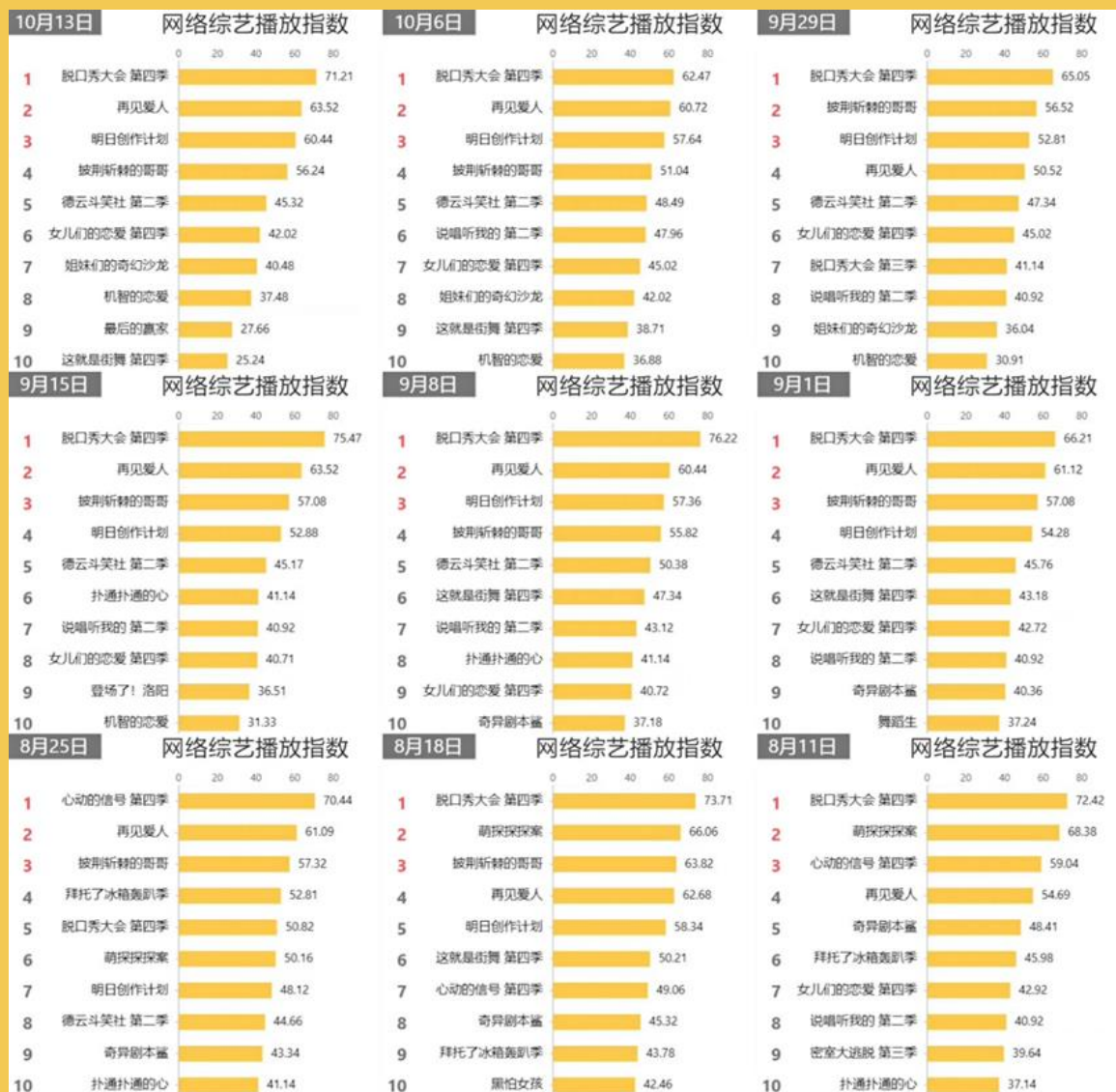


第一季和第二季的播放量都处于较低水平，但第一季有异常值
第三季和第四季的每期播放量都维持在较高水平，相比于前两集几乎翻倍

1.4 《脱口秀大会》节目分析

网络综艺播放指数对比：

除了8月25日的排名中，脱口秀大会第四季排在第5位，其余期数都是位列第一，说明该节目的热度很高。



1.4 《脱口秀大会》节目分析

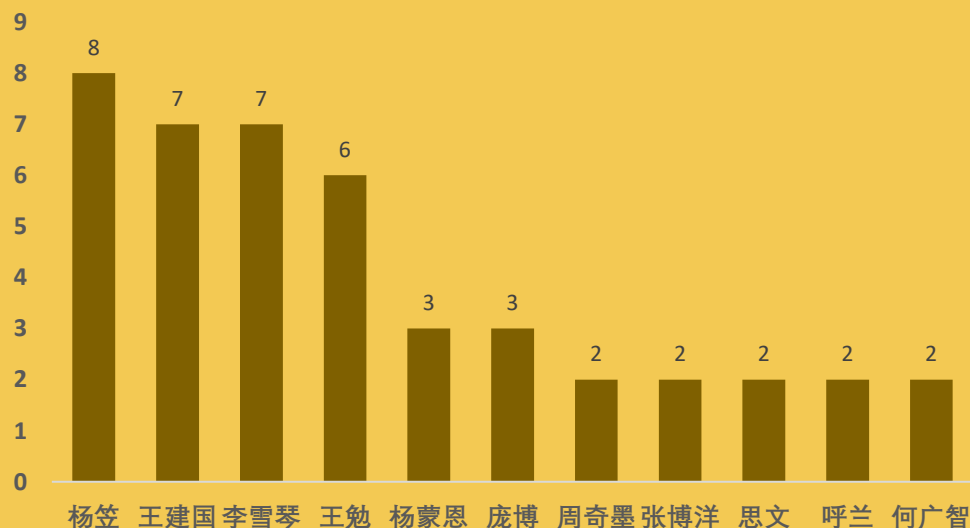
脱口秀演员分析：



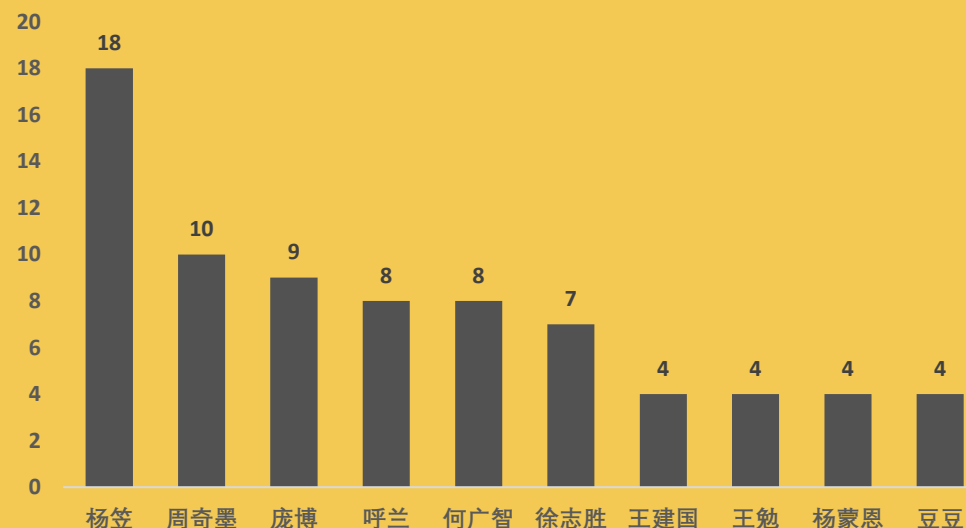
1.4 《脱口秀大会》 节目分析

第三季与第四季的热搜：

第三季热搜数



第四季热搜数



第三季平均每期上热搜7.2个，第四季平均每期上热搜11.3个

第三季的热搜主要来自杨笠、王建国、李雪琴等

第四季则来自于杨笠、周奇墨、庞博、呼兰、何广智等

两季中，杨笠的热搜数都是最高的

2

变量选择及构造

2.1 分析对象



本研究共选取**28**名脱口秀演员作为研究对象。



本季共56名参赛选手，但由于一轮游的选手可供分析的信息太少，所以最终选取第一轮晋级的28名选手作为分析对象。



每一名分析对象在该节目中至少有两场脱口秀表演可供分析。



2.2 因变量的选取及处理

选取——脱口秀演员的“火”

因变量	变量名称	来源	数据类型	数据说明
	中插代言数	《脱口秀大会四》	数值	节目中插代言个数
	微博粉丝数	微博	数值	微博粉丝数目
	搜索指数	百度	数值	百度搜索指数
	视频播放量	腾讯视频	数值	单人CUT播放量均值



微博粉丝数



中插代言数



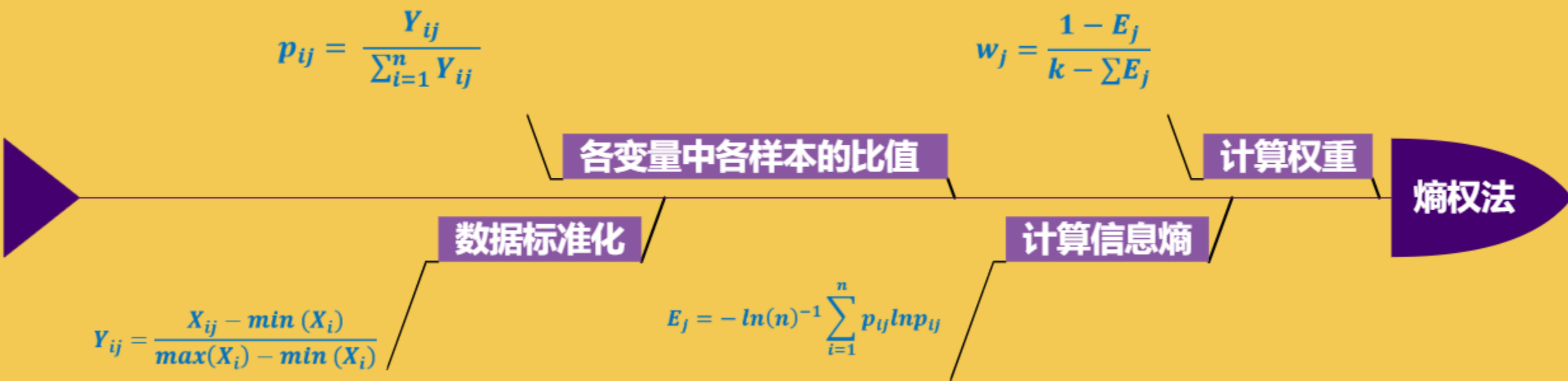
视频播放量



搜索结果数

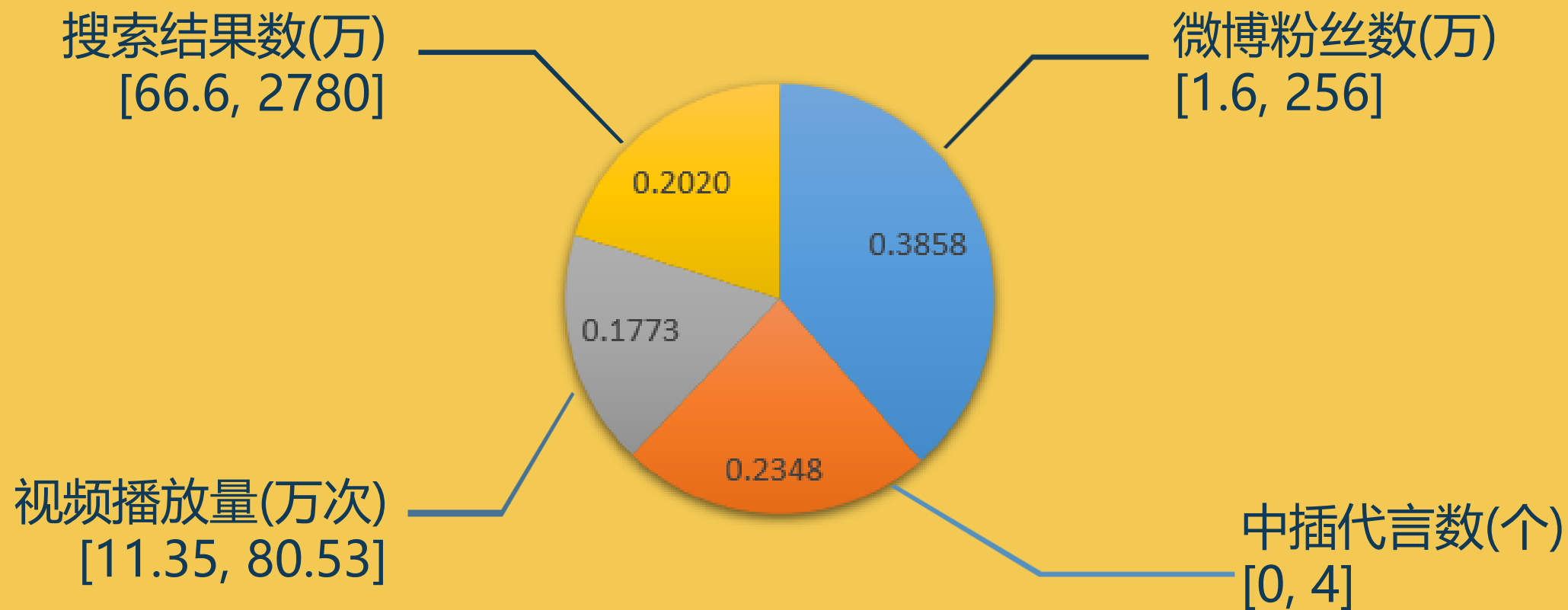
2.2 因变量的选取及处理

处理——熵权法建立综合评价体系



2.2 因变量的选取及处理

处理——各变量权重



2.3 自变量的选取

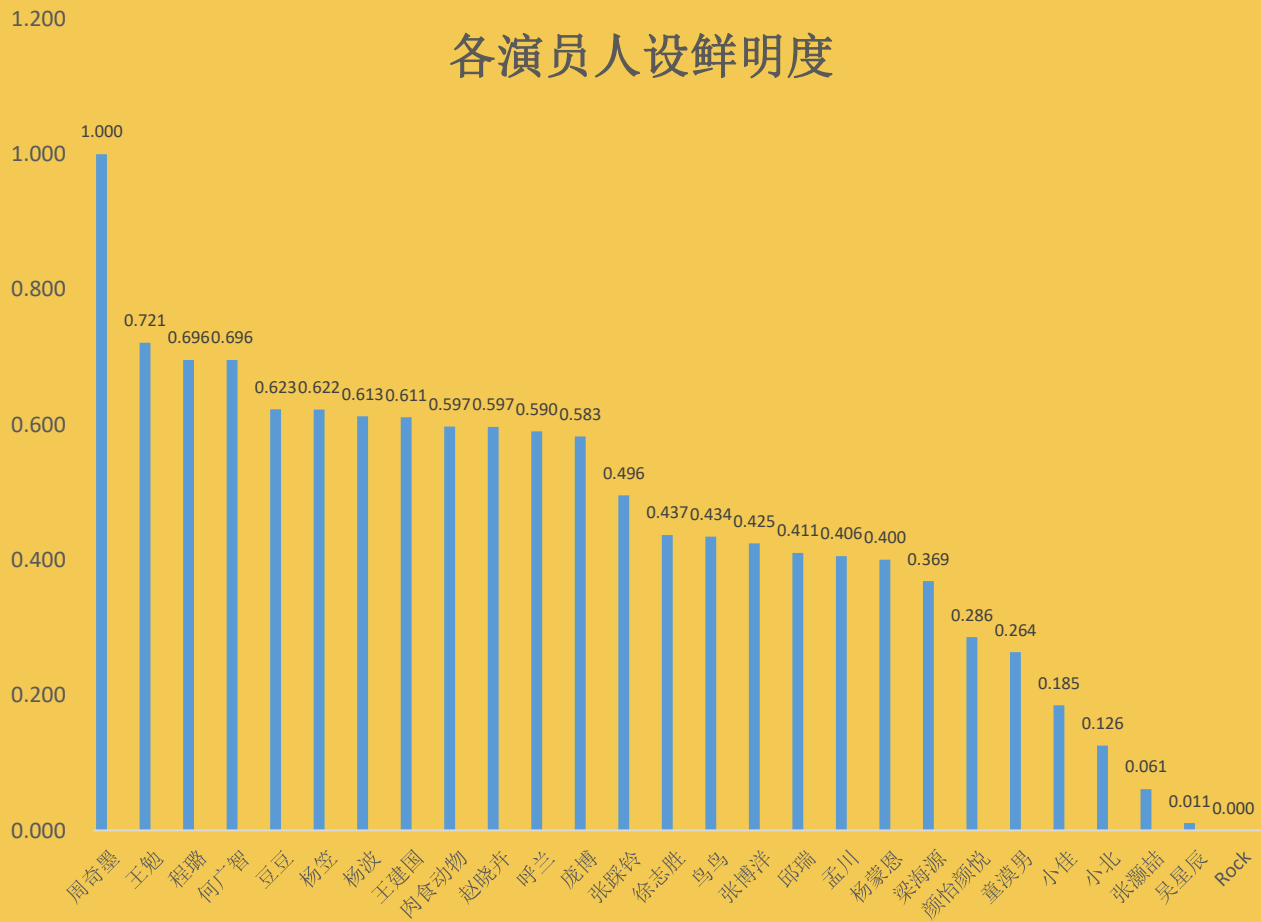
自变量	变量名称	来源	变量类型	数据说明
	表演年限	演员出场标签信息	数值	脱口秀表演年限
	出场期数	《脱口秀大会》	数值	各演员第四季参赛期数
	是否签约	演员微博身份标签、百度百科资料	二分类	是否与笑果文化签约
	是否组CP	微博超话、节目文案、评论等	二分类	在节目内是否与他人组CP
	是否参加过第三季	演员参赛信息	二分类	是否参加过第三季
	平均表演时长	《脱口秀大会》	数值	各期表演时长取均值
	文稿长度	《脱口秀大会》	数值	选取每个演员2-3篇表演文稿，求平均字数
	文稿情感得分	《脱口秀大会》	数值	用snowNLP对文稿逐句情感打分并求均值
	相关评论数	腾讯视频&豆瓣评论	数值	爬取3万多条评论，筛选演员相关评论 21290 条，计算各演员相关评论
	人设鲜明度	腾讯视频&豆瓣评论	数值	人设关键词出现频次占比
	观众认可度	腾讯视频&豆瓣评论	数值	对各演员相关评论进行情感打分并求均值

2.4 自变量描述性分析

构造&描述——人设鲜明度(10/30)



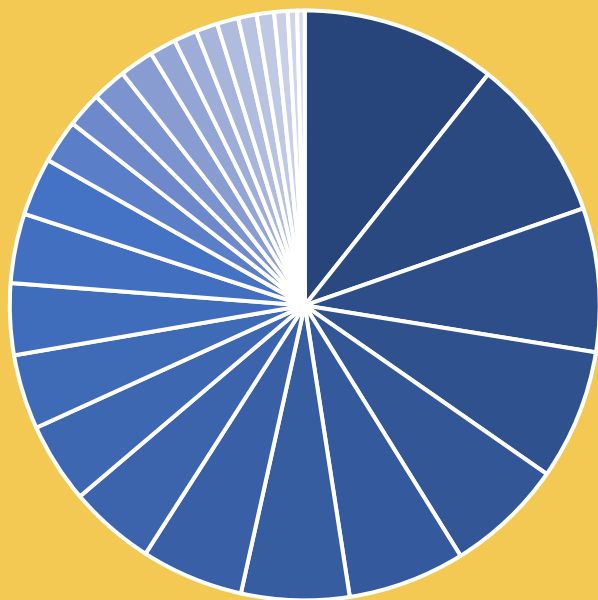
各演员人设鲜明度



2.4 自变量描述性分析

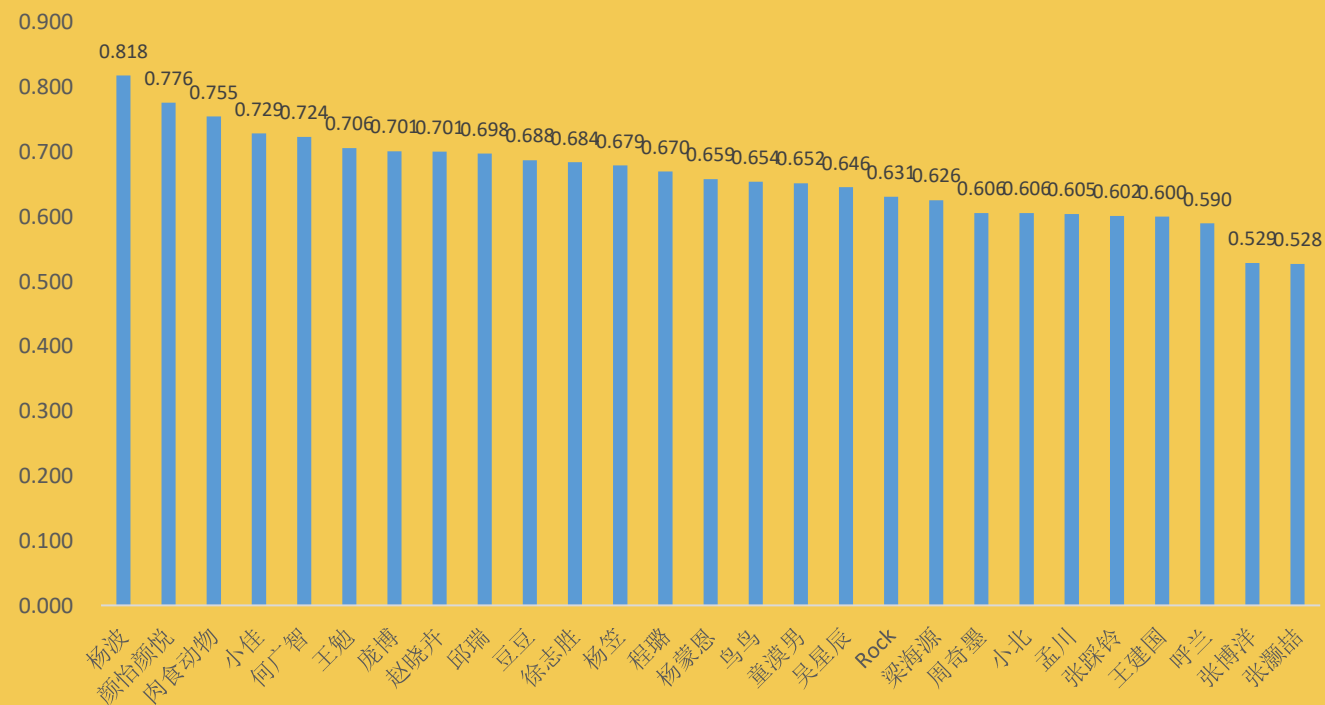
描述——评论占比&观众认可度

各演员相关评论数占比



呼兰 王勉 肉食动物 张踩铃 杨笠 邱瑞 张博洋 赵晓卉 周奇墨 杨蒙恩 童漠男 小佳 豆豆 王建国 颜怡颜悦 Rock 庞博 杨波 程璐 吴星辰 何广智 鸟鸟 梁海源 徐志胜 孟川 小北

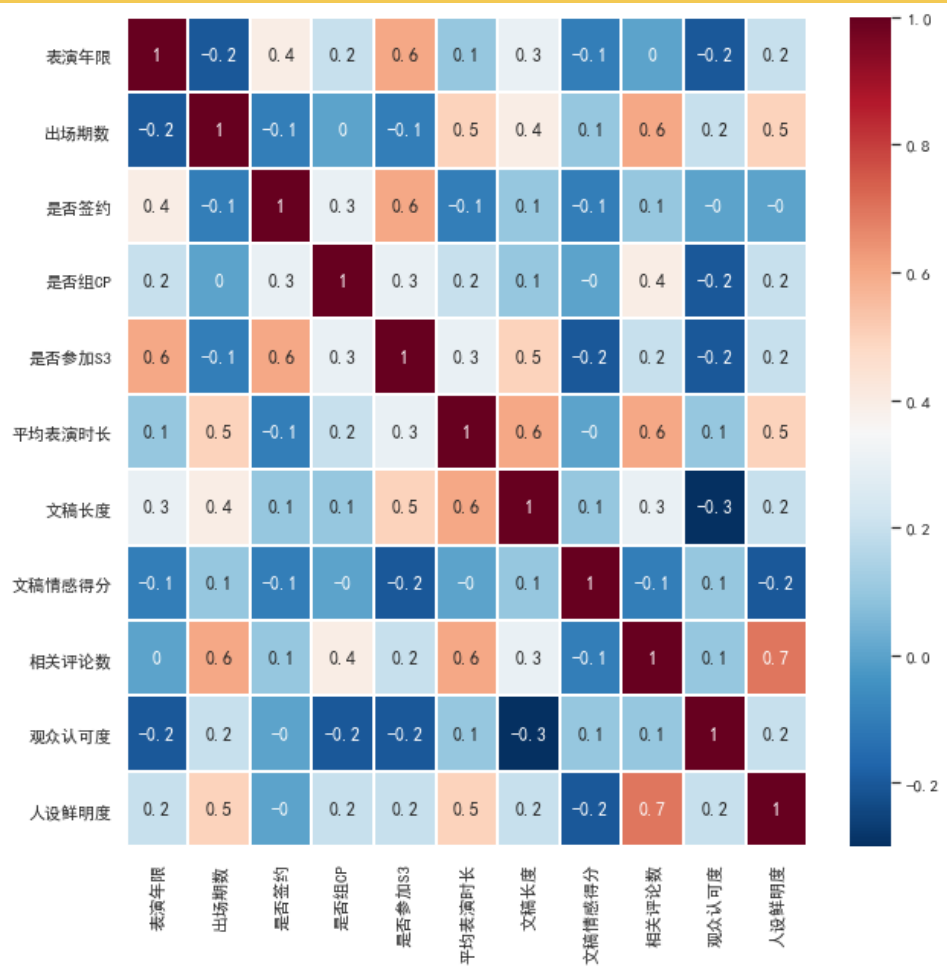
观众平均认可度



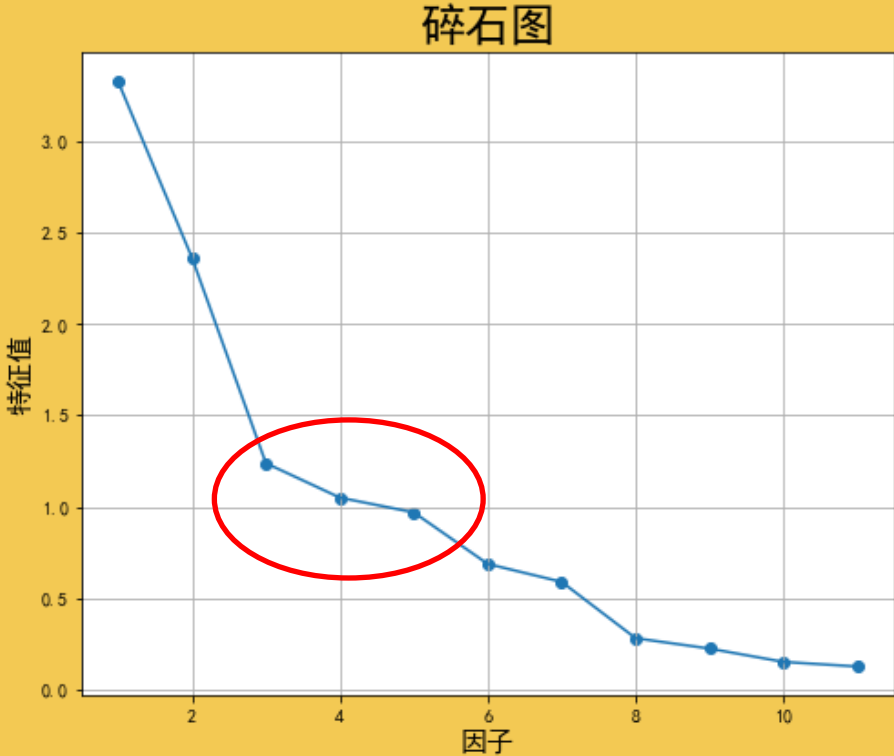
3

建模分析

3.1 因子分析



KMO检验和Bartlett的检验	
KMO	0.5403
Bartlett球形检验	2.5707e-07

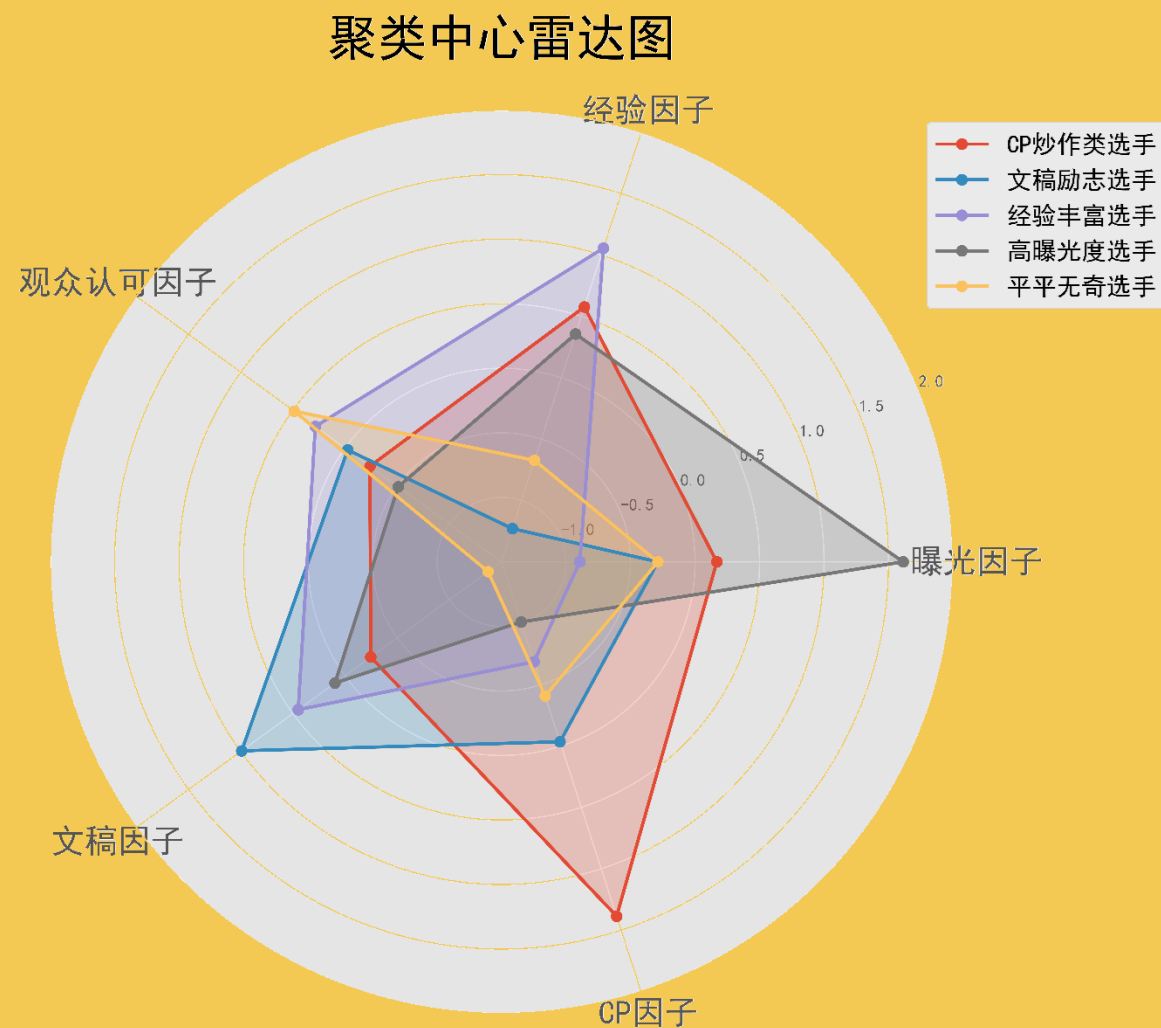


因子载荷

	曝光因子	经验因子	观众认可因子	文稿因子	CP因子
表演年限	0.0099	0.8095	-0.1347	-0.1056	0.0000
出场期数	0.7957	-0.1743	0.1396	0.1367	-0.0316
是否签约	-0.1431	0.7868	0.1852	0.0687	0.3226
是否组CP	0.1523	0.2061	-0.1368	0.0102	0.9089
是否参加S3	0.1958	0.8579	-0.1937	-0.1312	0.0597
表演时长	0.8483	0.0876	-0.1313	0.0216	0.0264
文稿长度	0.6186	0.3481	-0.5534	0.2297	-0.2205
文稿情感得分	-0.0001	-0.1168	0.0713	0.9401	0.0114
相关评论数	0.8141	0.0387	0.0748	-0.0716	0.4335
观众认可度	0.1941	-0.0822	0.8829	0.1134	-0.1658
人设鲜明度	0.7276	0.1275	0.2789	-0.3466	0.0769

根据因子得分聚类

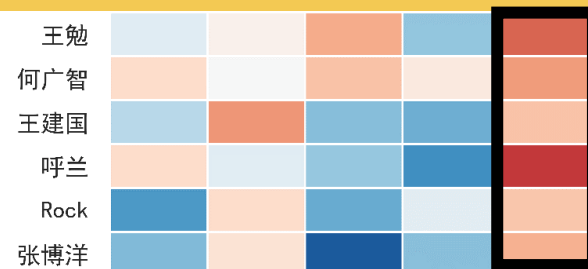
- 根据因子得分，对所有演员做Kmeans聚类
- 根据手肘法，取K=5
- 聚类中心的雷达图如右图所示



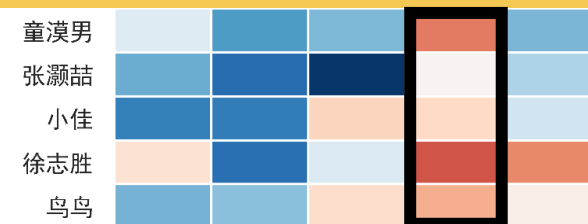
因子得分聚类



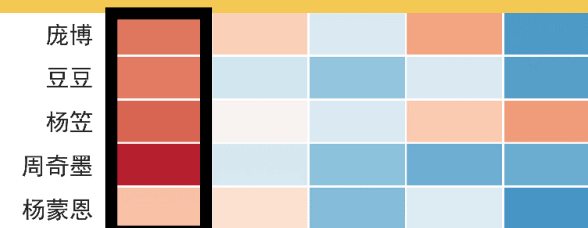
CP炒作类选手



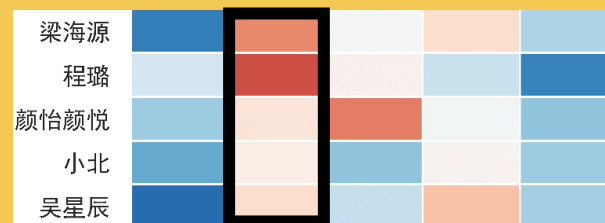
文稿励志选手



高曝光度选手



经验丰富选手



曝光因子

经验因子

观众认可因子

文稿因子

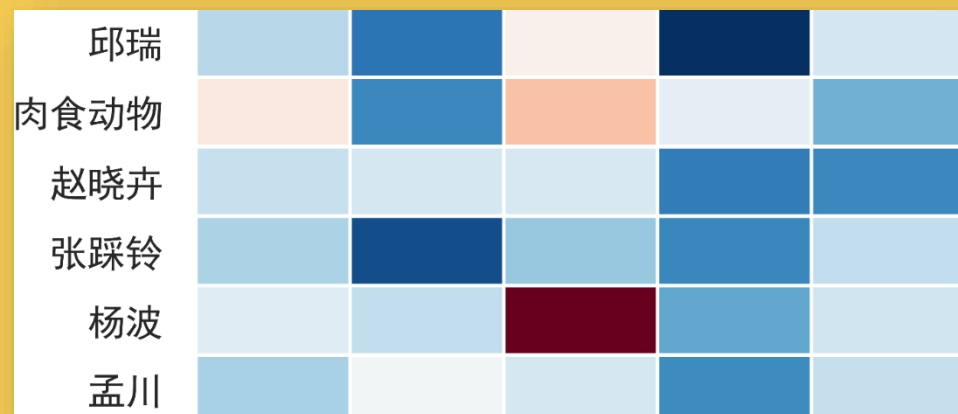
CP因子



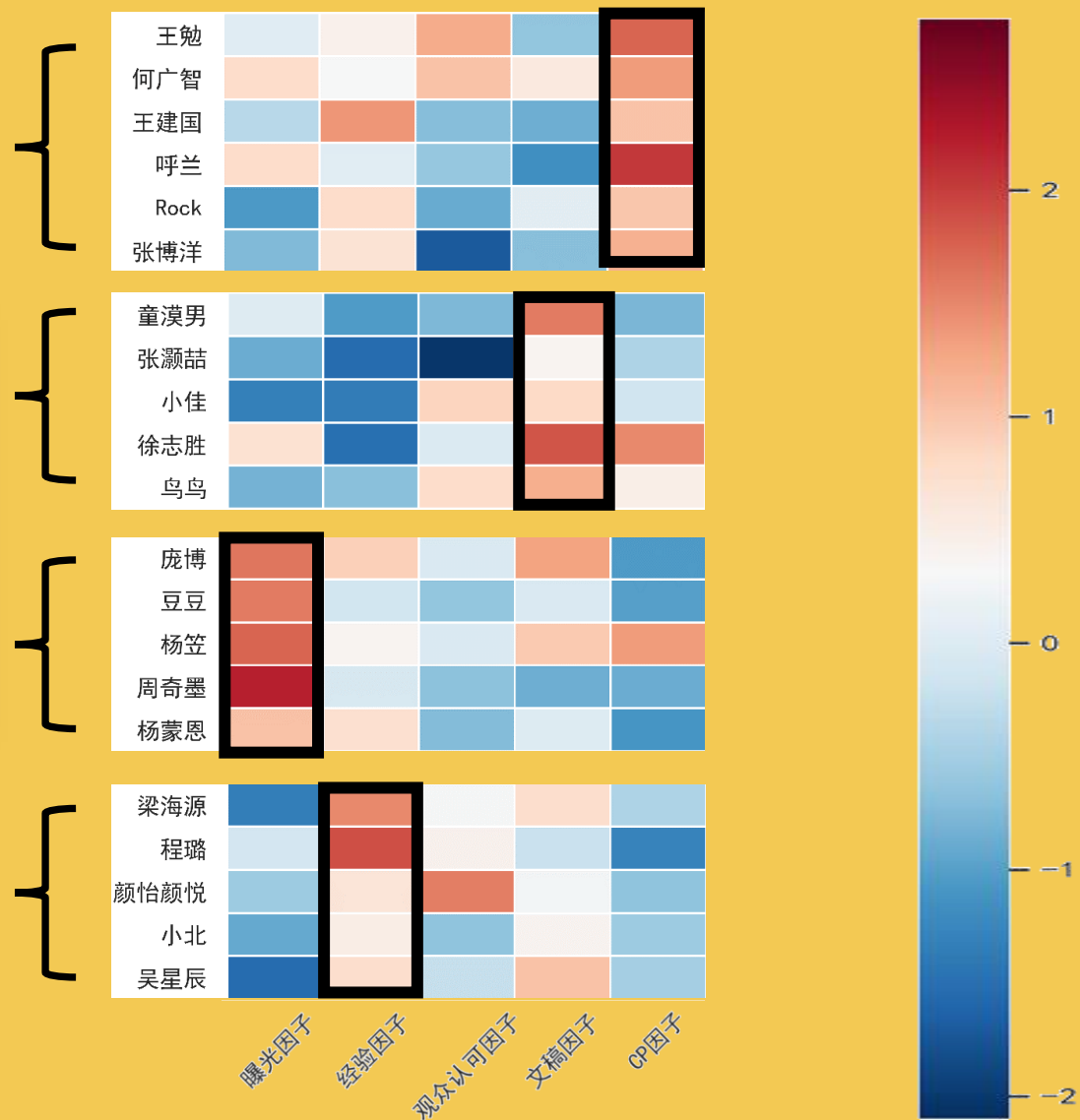
因子得分聚类



CP炒作类选手



经验丰富选手



因子得分聚类

聚类中心	曝光因子	经验因子	观众认可因子	文稿因子	CP因子	代表人物	平均综合热度
1. CP炒作类选手	0.1690	0.5770	-0.2390	-0.2460	1.3870	何广智	0.4634
2. 文稿励志选手	-0.2880	-1.2290	-0.0250	0.9910	-0.0340	小佳	0.1848
3. 经验丰富选手	-0.8930	1.0550	0.2850	0.4490	-0.6860	程璐	0.2418
4. 高曝光度选手	1.6140	0.3560	-0.5100	0.0970	-1.0100	庞博	0.5091
5. 平平无奇选手	-0.2880	-0.6730	0.4870	-1.3710	-0.4070	赵晓卉	0.1773

- CP 炒作类选手和高曝光度选手的综合热度最高
- 平平无奇选手的综合热度最低

4

结论分析

4.1 总结

以下方法可能有助于提高脱口秀演员的热度

- 可与其他演员组CP来获取观众的关注，如南广智，北志胜
- 保持鲜明的人设，提高辨识度，如程璐的领导人设
- 文稿长度/段子数越多，演员越受欢迎，如周奇墨、杨波
- 犀利的文稿或新鲜的段子或新奇的表演可能更容易吸引观众，如杨笠、豆豆、肉食动物

4.2 不足

- 本案例分析中对于样本量的选取较小，数据分析结论的可推广性较低
- 对于观众评论的处理尚有不足，未能区分同一条评论中对不同演员的评价内容
- 演员文稿收集的数量较少，且对文稿的处理相对粗糙，不足以充分提取信息以表现文稿对于脱口秀演员的重要性



感谢观看!