

目录CONTENT

01 数据介绍

- bilibili_data.csv数据集包含信息及变量的基本含义

02 Up主画像

- Up主个人信息
- Up主主页信息

03 变量间关系

- 数据处理与相关系数
- 一对一分析

04 建模分析

- 岭回归介绍
- 建模与拟合结果

05 结论

- 创作生态
- 成长建议

数据相关性 (宏观)

数据处理

相关系数与初步可视化

/03



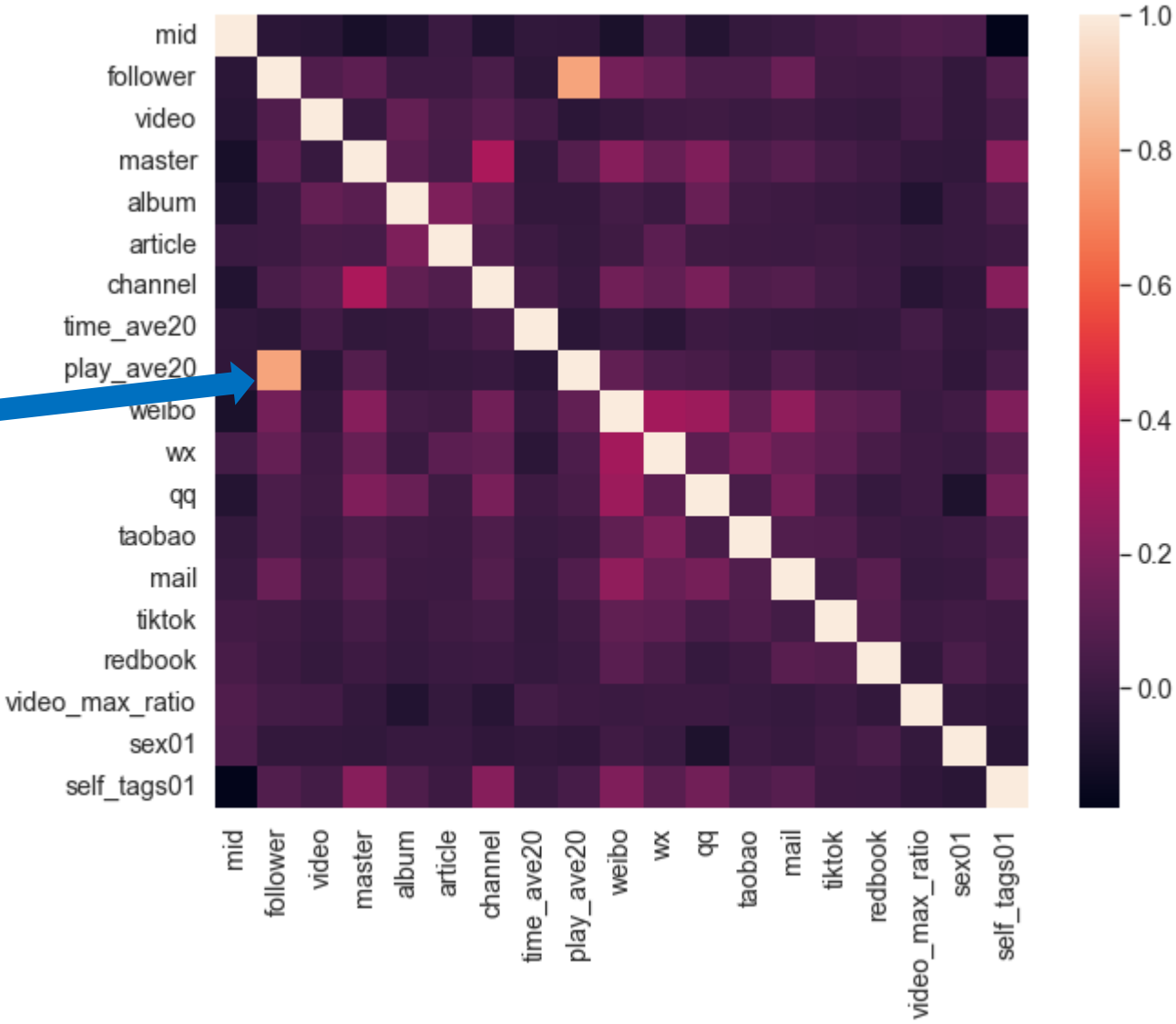
数据处理

将离散变量转为0/1变量

sex	self_tags	sex01	self_tags01
男	['拖稿大师', '上古老坟', '丸子', '话唠', '绅士']	0.0	1
女	[]	1.0	0
男	[]	0.0	0
男	['猫和老鼠', '真人快打', 'MMD', '兔八哥', '高飞狗']	0.0	1
保密	['虚拟主播', '搬运', '翻译']	0.5	1
...
男	[]	0.0	0
保密	[]	0.5	0
保密	[]	0.5	0
保密	[]	0.5	0
保密	[]	0.5	0

相关性 (系数)

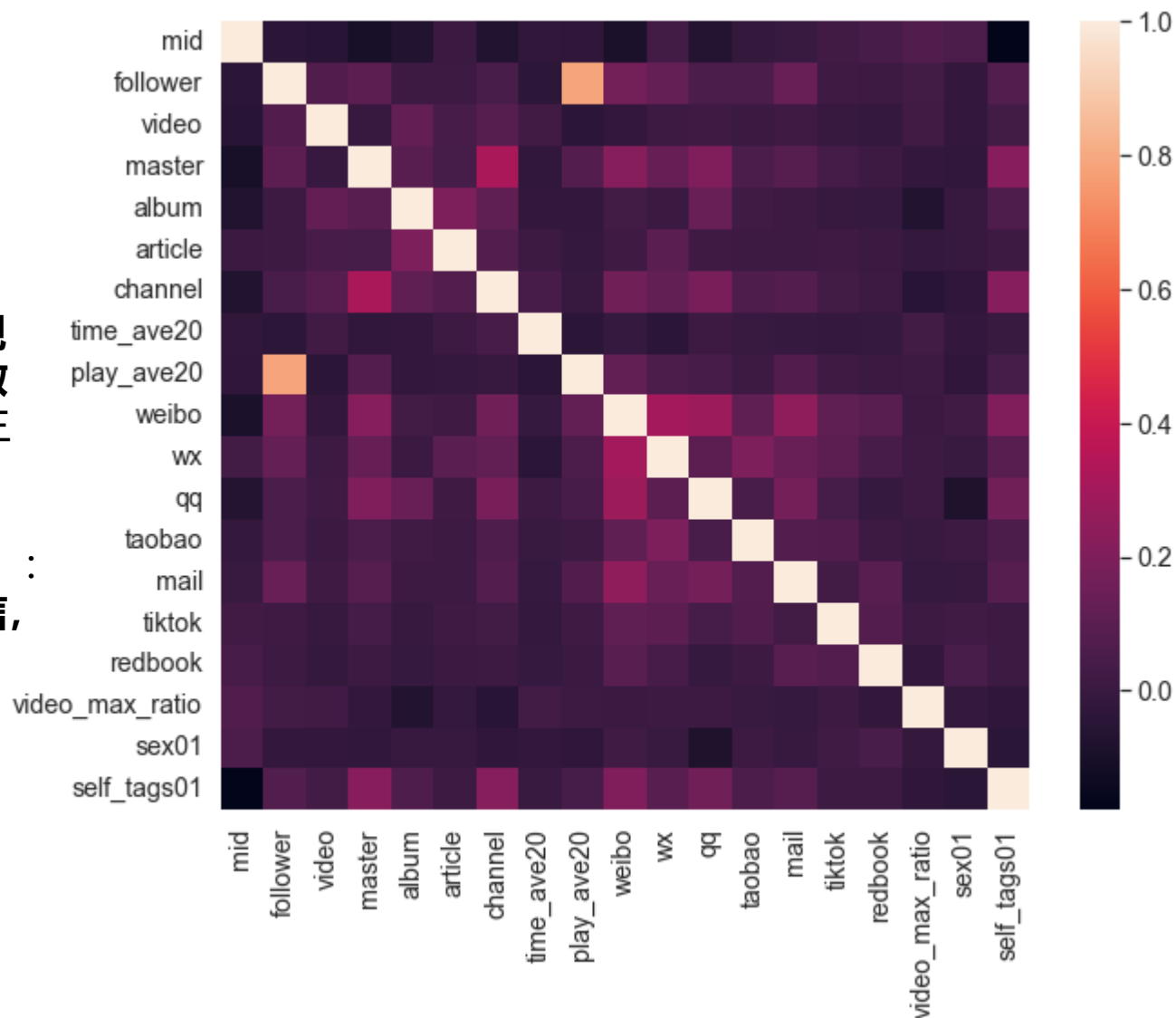
近期视频平均播放量
与粉丝数



相关性 (系数)

观察发现**粉丝数 (follower)**和**近期20个视频的平均播放量**相关系数极大, 以及**频道数**与**代表作个数**相关性较大, **微博 微信 QQ**三者相关性较大

与**粉丝数**相关的 (相关系数绝对值大于0.1) :
代表作个数, 近期平均播放量, 微博, 微信, 邮箱

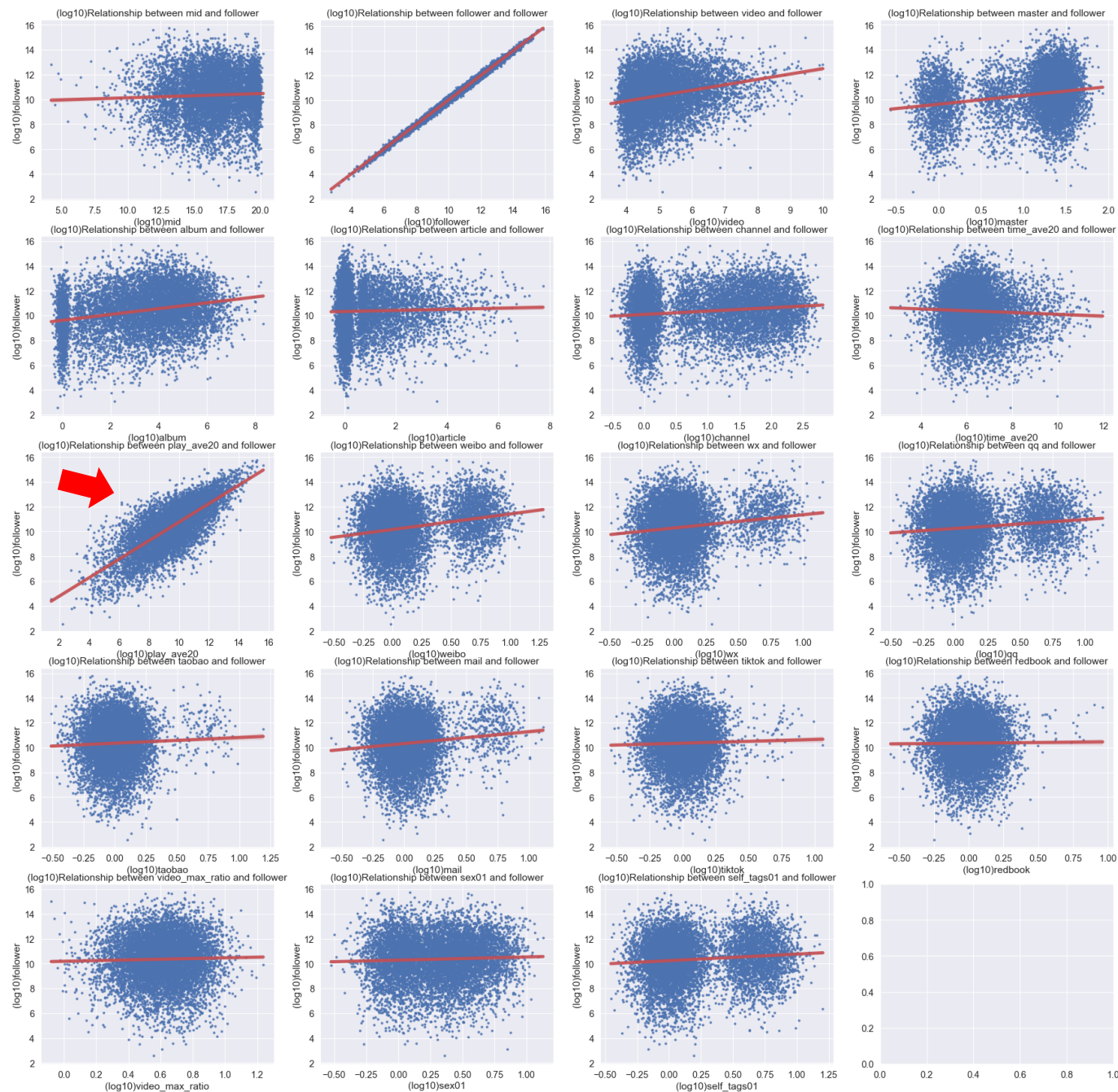


与粉丝数相关

相关性 (scatter)

与粉丝数相关

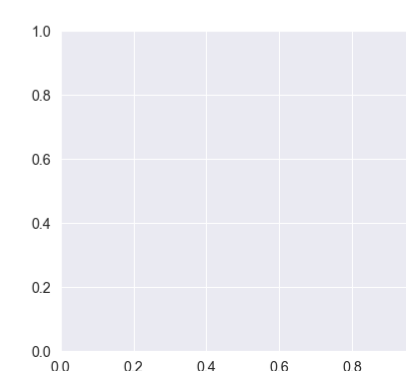
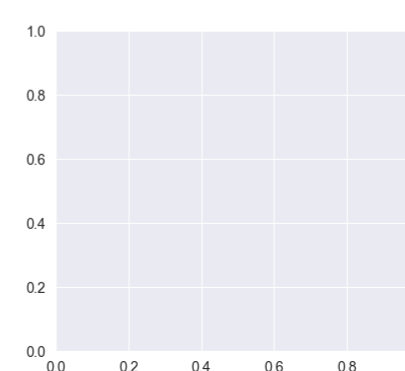
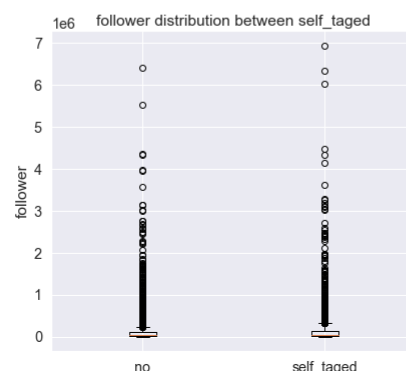
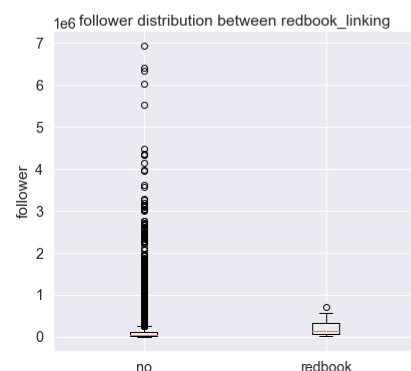
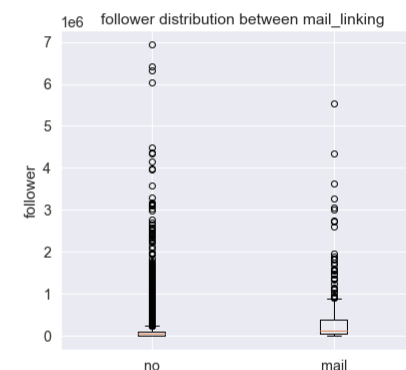
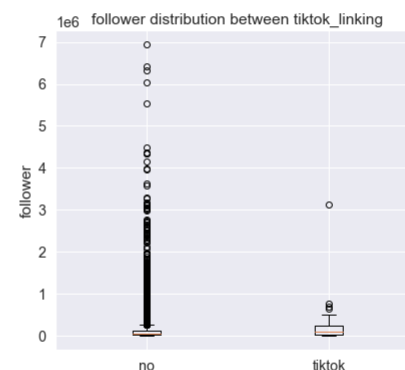
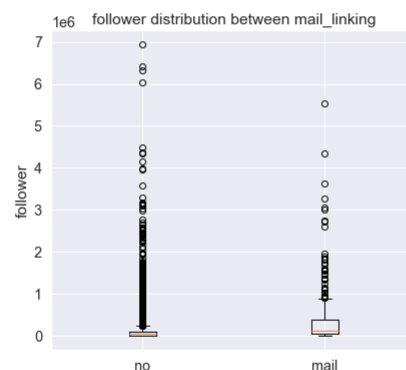
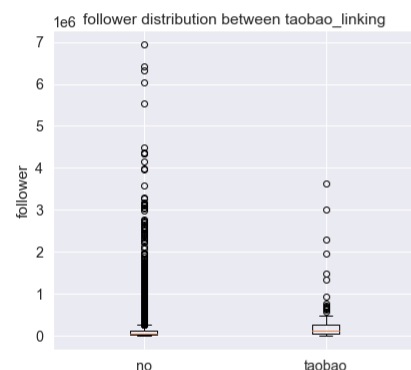
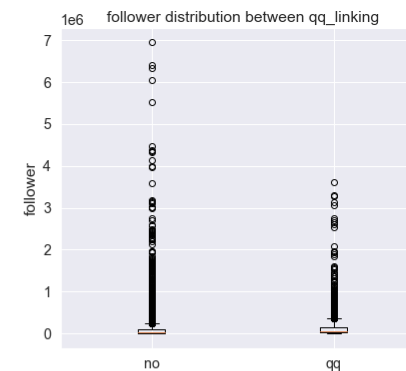
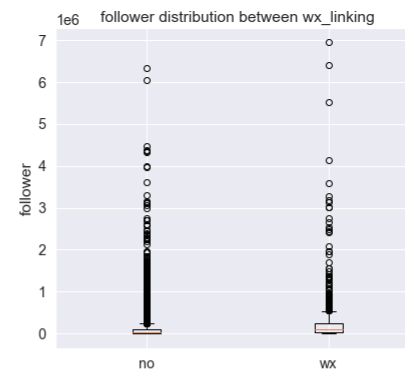
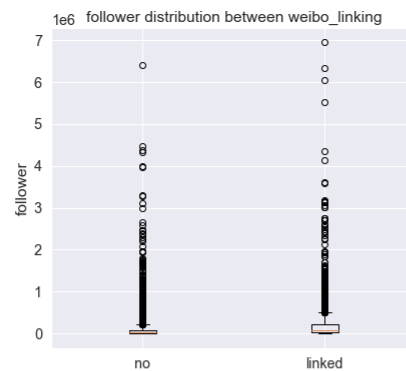
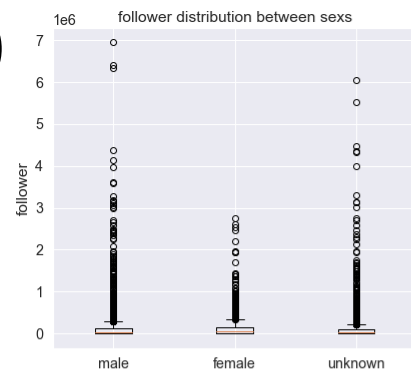
取log后水平打散plot



相关性 (scatter)

与粉丝数相关

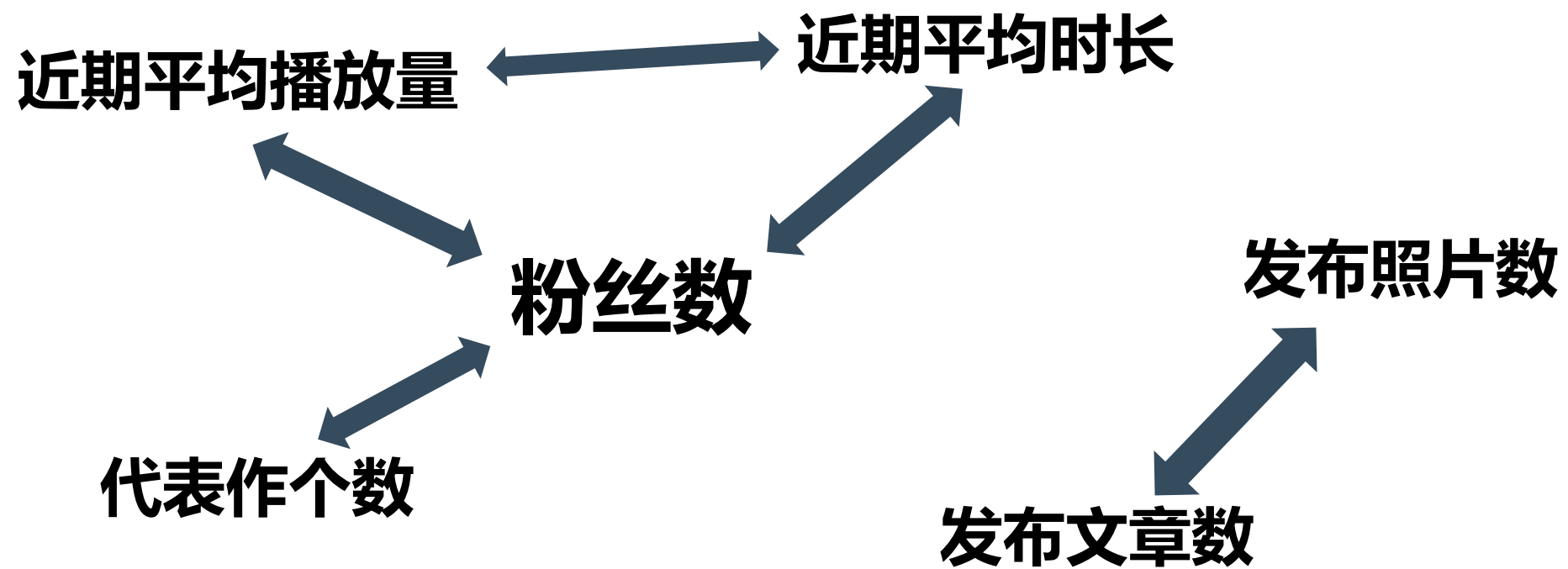
离散变量boxplot



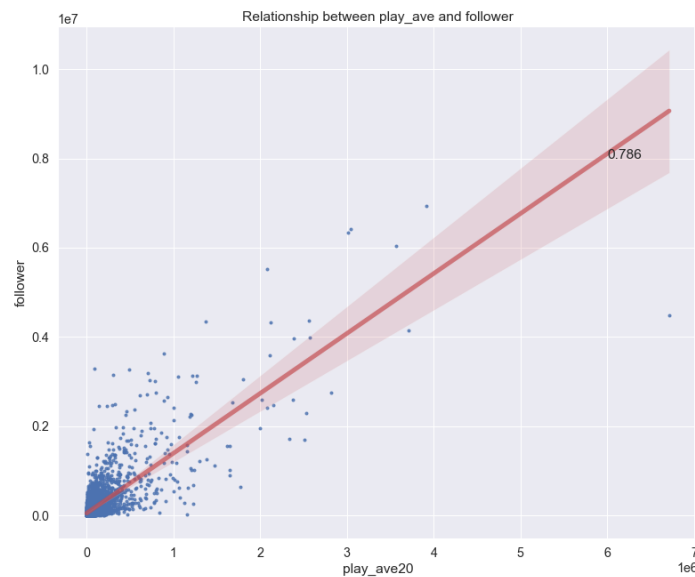
变量相关性（一对一）

/03

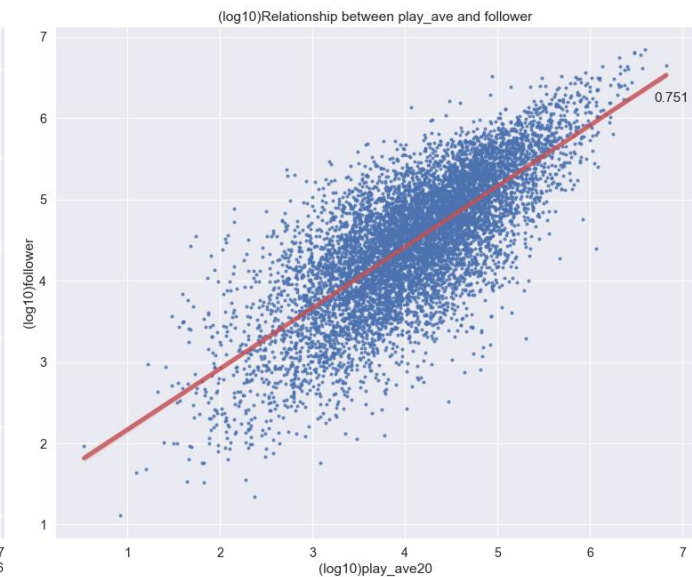




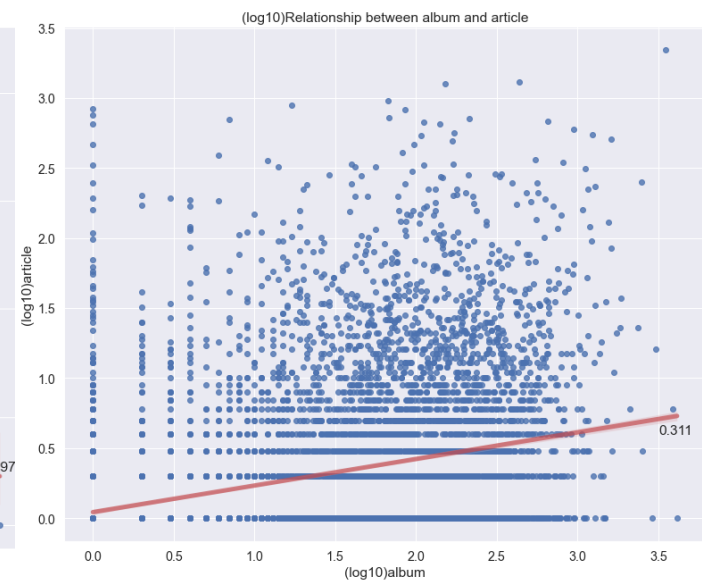
近期平均播放量 与 粉丝数



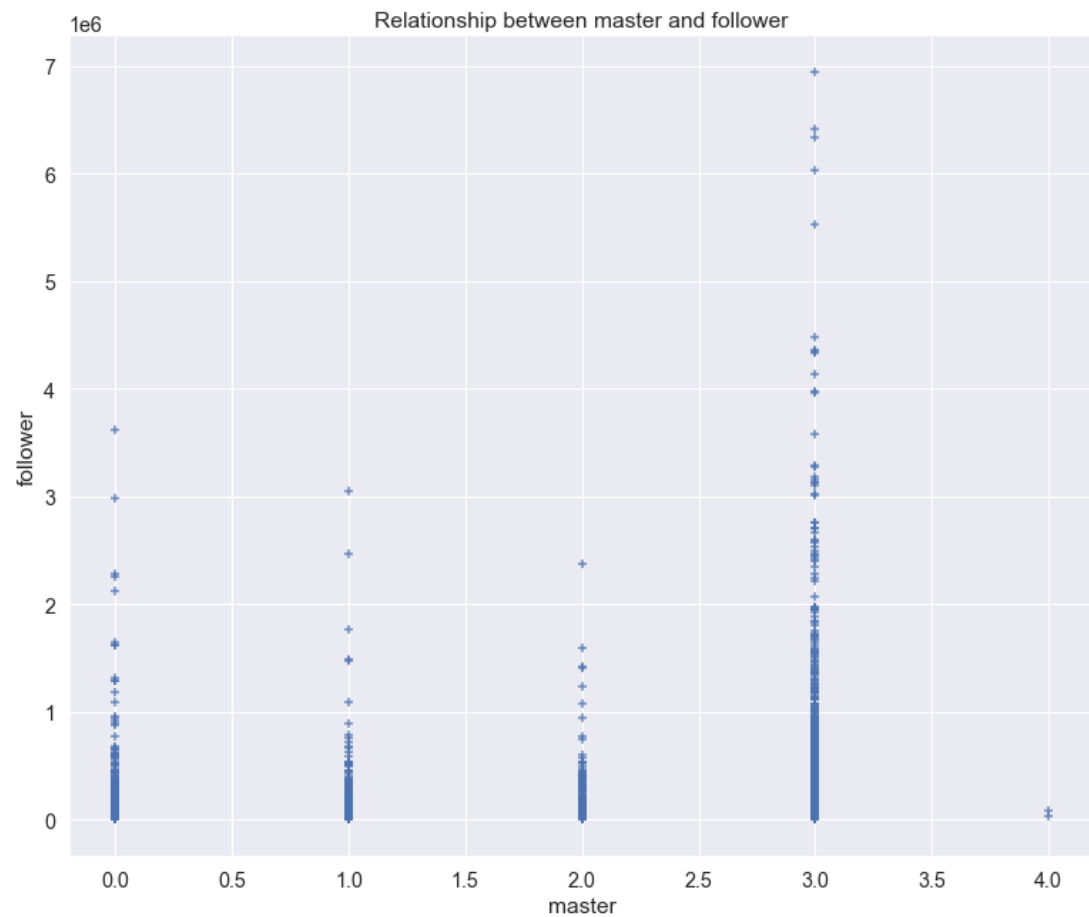
Log变换后



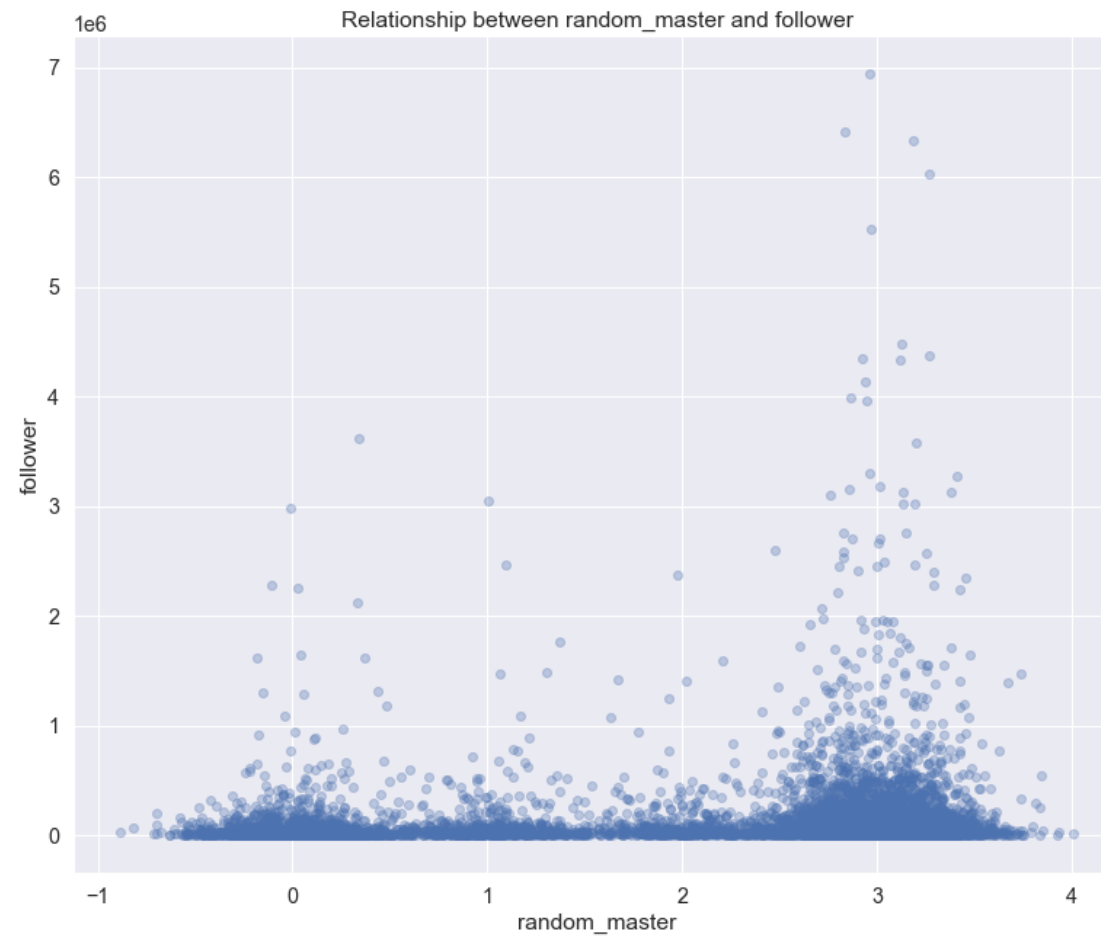
发布文章数 与 照片数



代表作个数 与 粉丝数

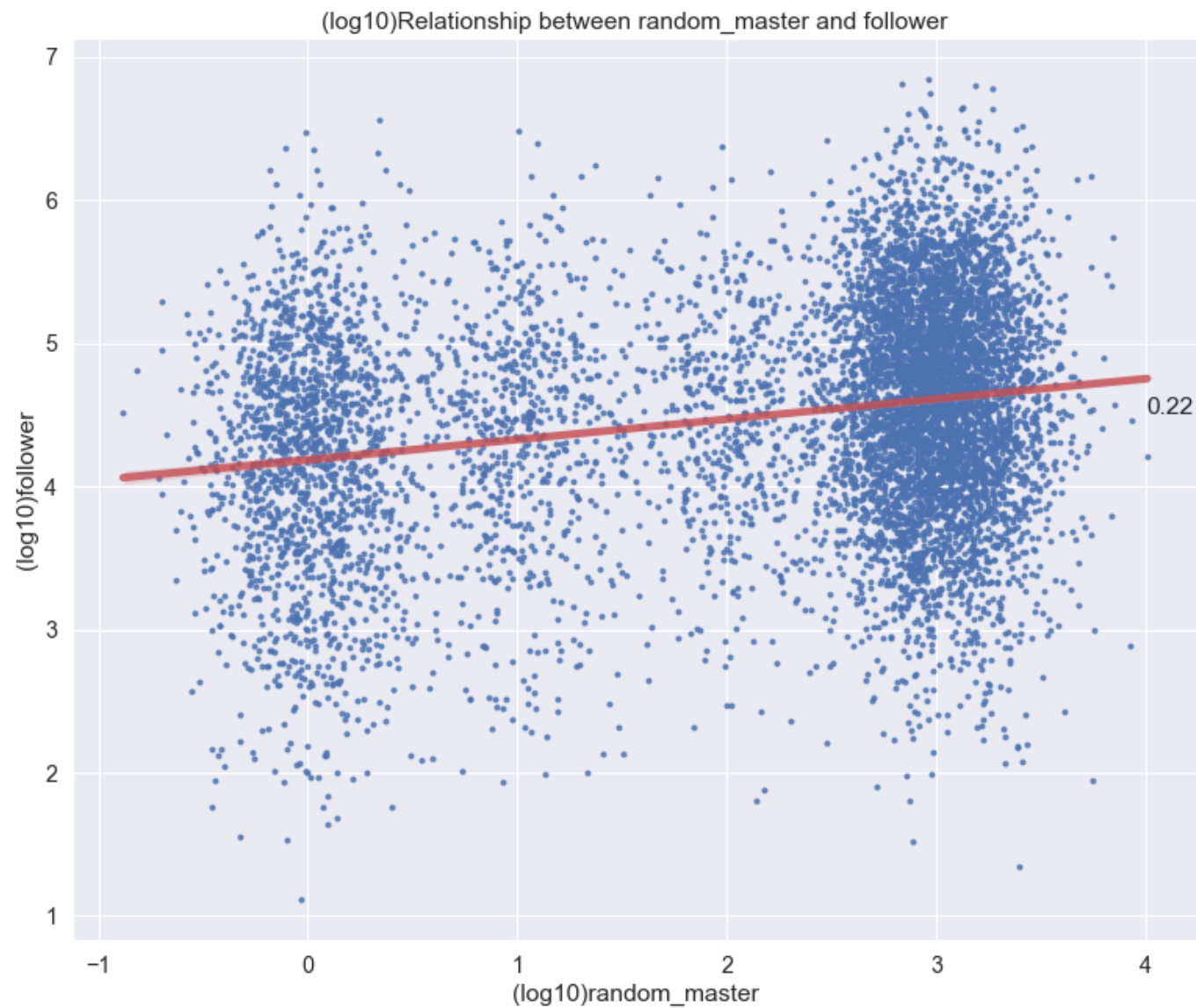


水平打散

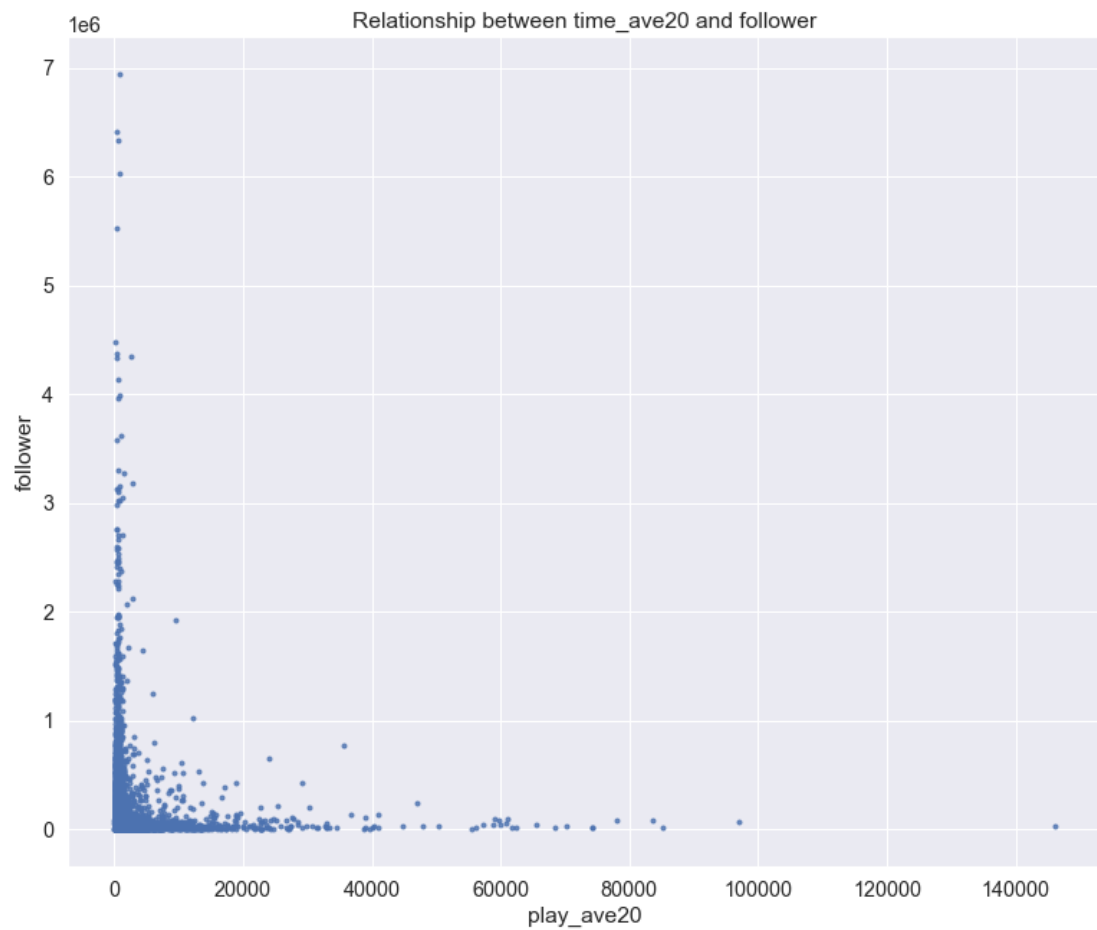


代表作个数 与 粉丝数

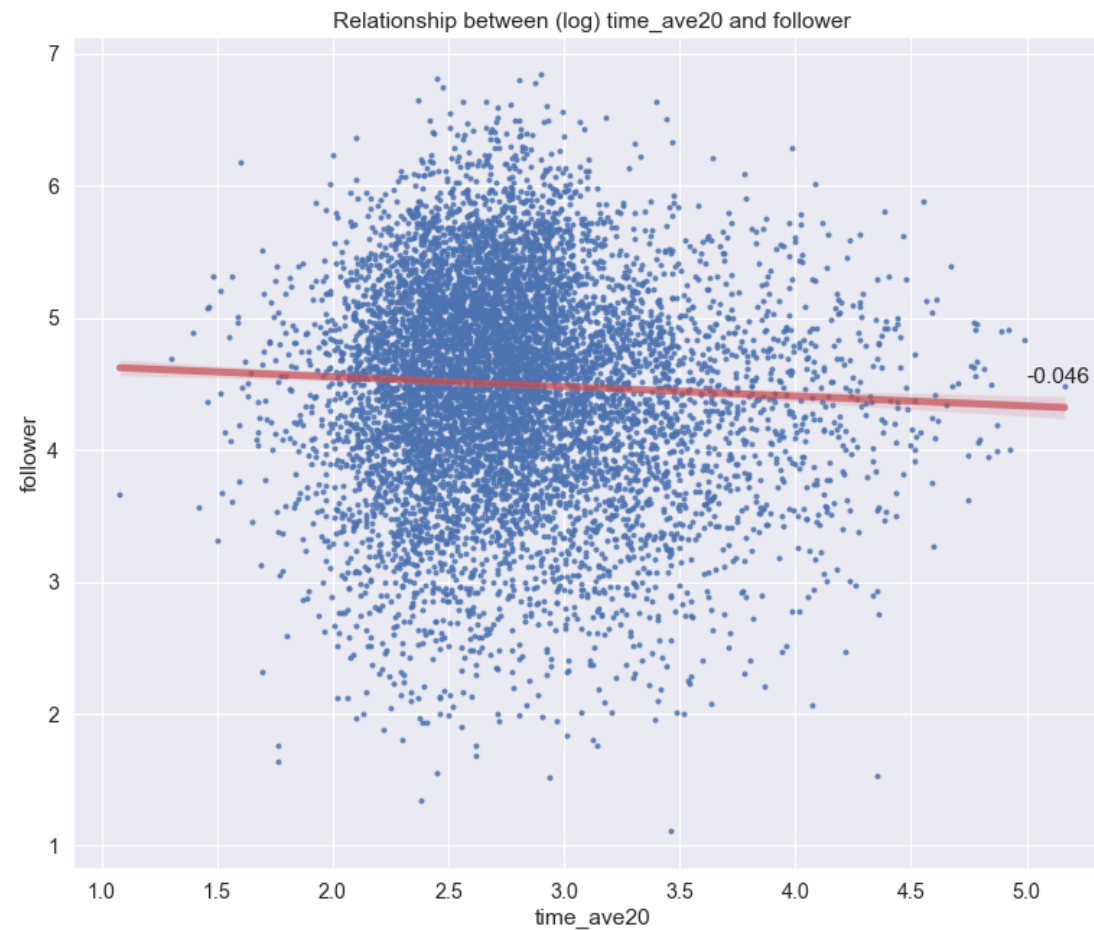
Log变换后



近期平均视频时长 与 粉丝数



Log变换后

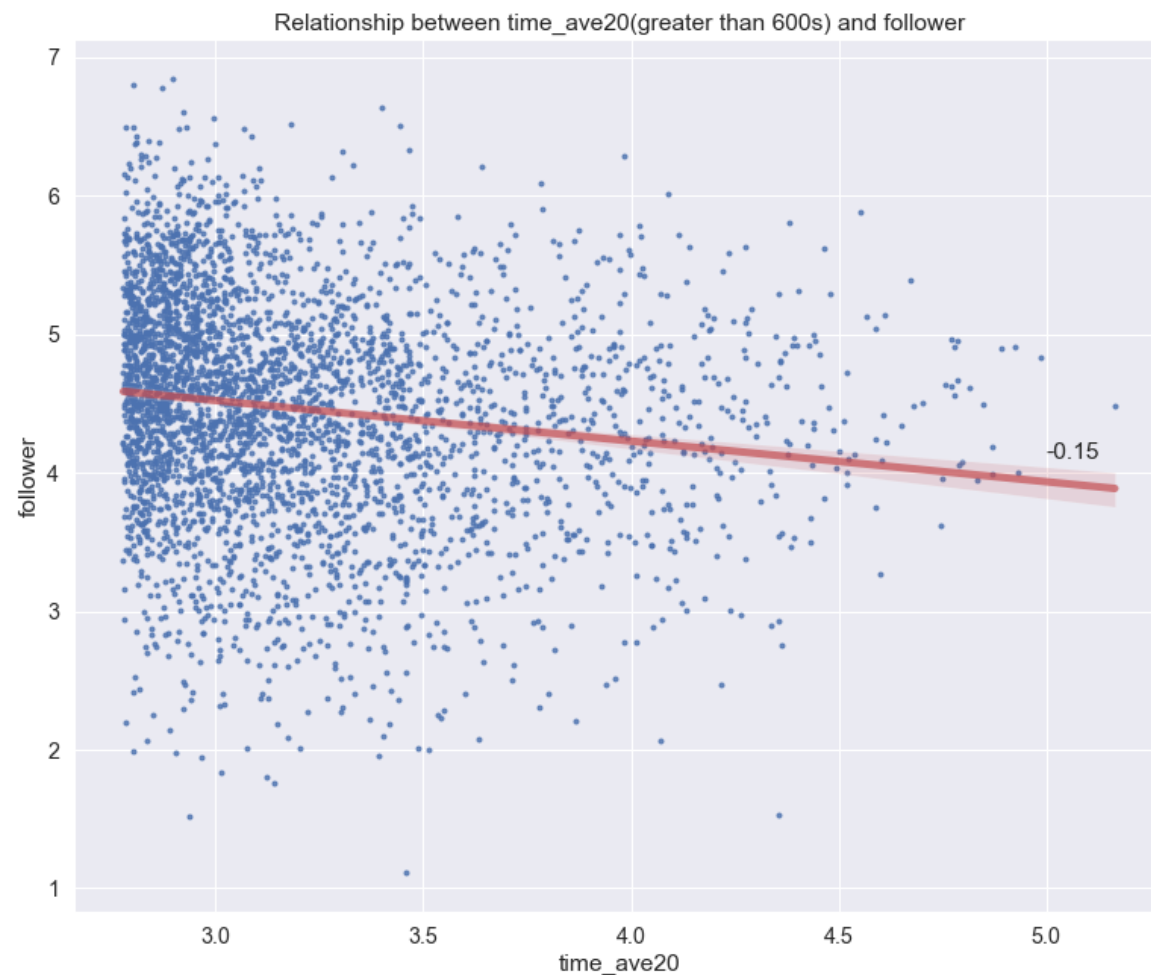


近期平均视频时长 与 粉丝数

小于10分钟

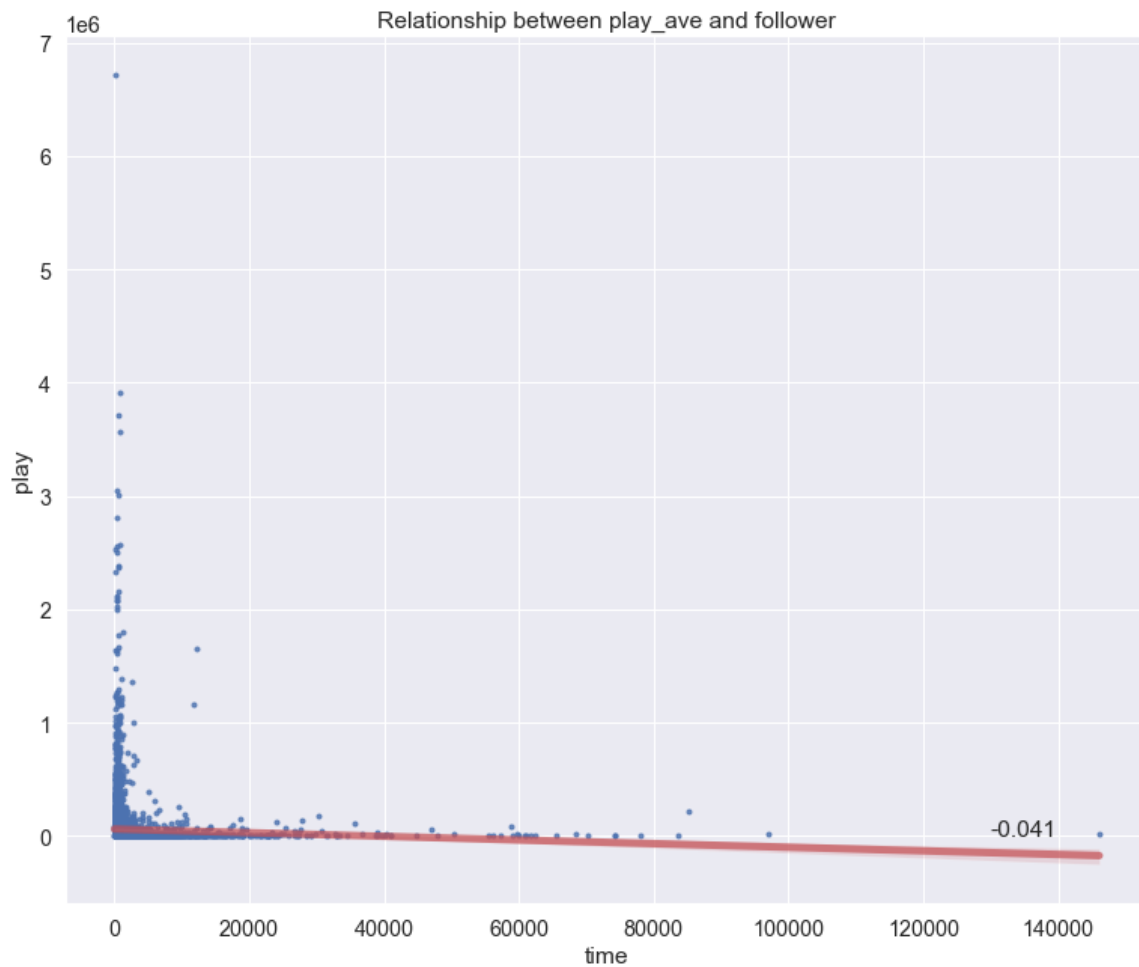


大于10分钟



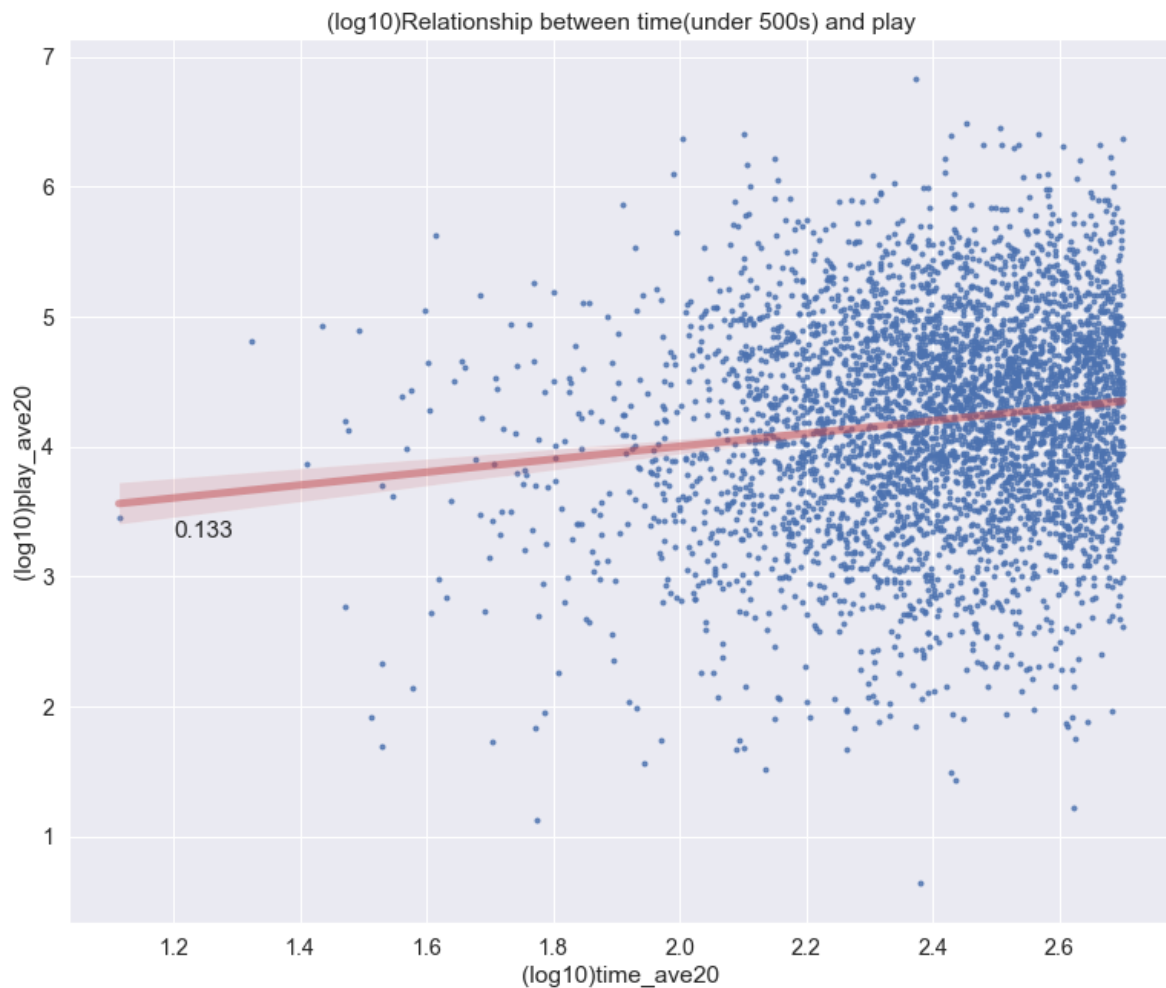
近期视频平均时长 与 近期视频平均播放量

Log变换后



近期视频平均时长 与 近期视频平均播放量

小于500s



大于500s

