

抖音大 V 分析

数据分析报告

Young

目录

- 背景介绍 2
- 问题介绍 2
- 程序设计思路 3
- 结果与分析 3
 - 粉丝增长量折线图 3
 - 粉丝总量折线图 4
 - 多元线性回归分析 4
 - 文件说明 5

背景介绍

抖音是一个面向全年龄的音乐短视频社区平台，如今已成为最火的短视频软件，无数短视频创作者通过抖音分享生活，分享技能，分享美好。其中有点赞、评论、转发、关注等功能。

我们试图分析“大 V”即粉丝数量相对较多的博主的涨粉数量和点赞、评论、转发之间的关系，并试图量化。

问题介绍

我们试图通过分析抖音大 V：“人民日报”的 2021 年 3 月份的数据，通过**多元线性回归**得出相关关系。我们得到了“人民日报 3 月粉丝增长数据.csv”和“人民日报 3 月份视频数据.csv”两份数据，其中包含了博主“人民日报”从 2021 年 3 月 1 日到 2021 年 3 月 31 日的粉丝数量变化和每一条视频的点赞、评论、转发数量。此外还得知在 3 月 1 日，其粉丝数量为 122747343 个。

我们有以下任务：

- 作出粉丝增长量以及粉丝总量折线图。
- 分析涨粉与当天点赞，评论，分享量的关系。使用多元线性回归，求出最佳拟合线的截距 d 以及回归系数 (α, β, γ)

程序设计思路

使用 Python 的版本为 3.7, IDE 为 PyCharm Community Edition 2019。

通过 Python 以及相关库进行处理。

具体为：通过 csv 库进行 csv 数据的导入和导出，

通过 statsmodels 库进行多元线性回归，

通过 matplotlib 库进行绘图，

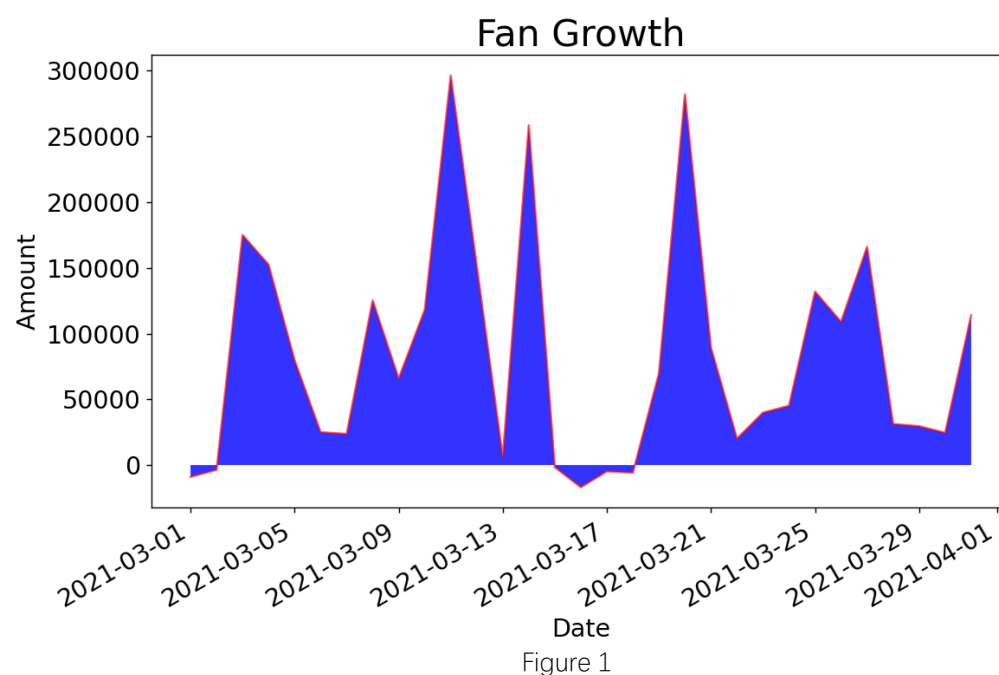
以及通过 numpy、pandas、datetime 库进行相关数据处理。

具体代码实现见附带代码。

结果与分析

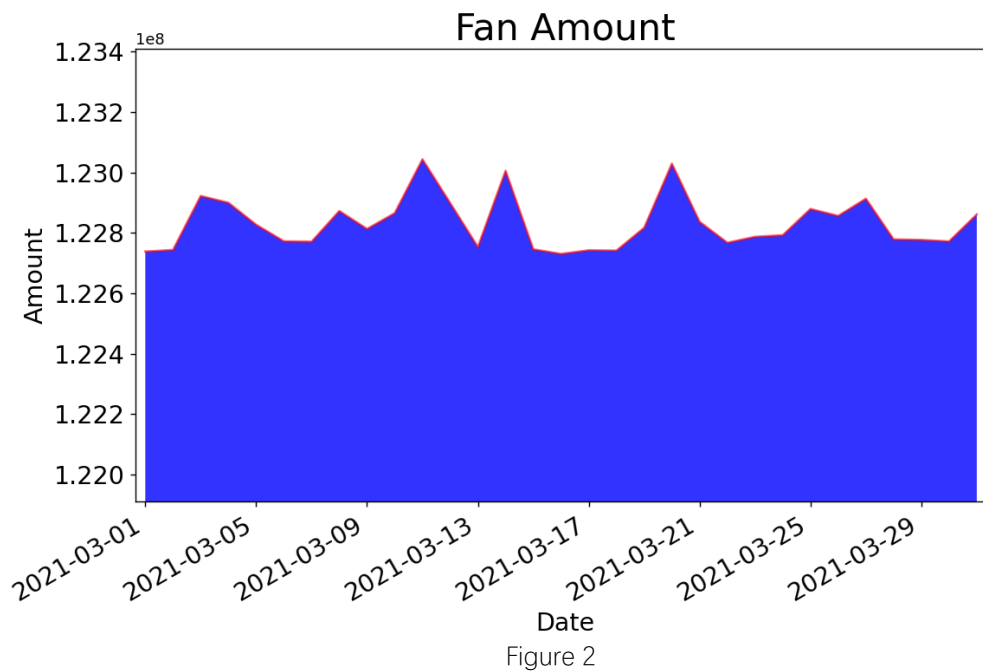
粉丝增长量折线图

见 Figure 1



粉丝总量折线图

见 Figure 2



多元线性回归分析

使用 OLS 回归模型，程序输出结果如下：

截距：

$$d = 3.64713375e + 04$$

系数：

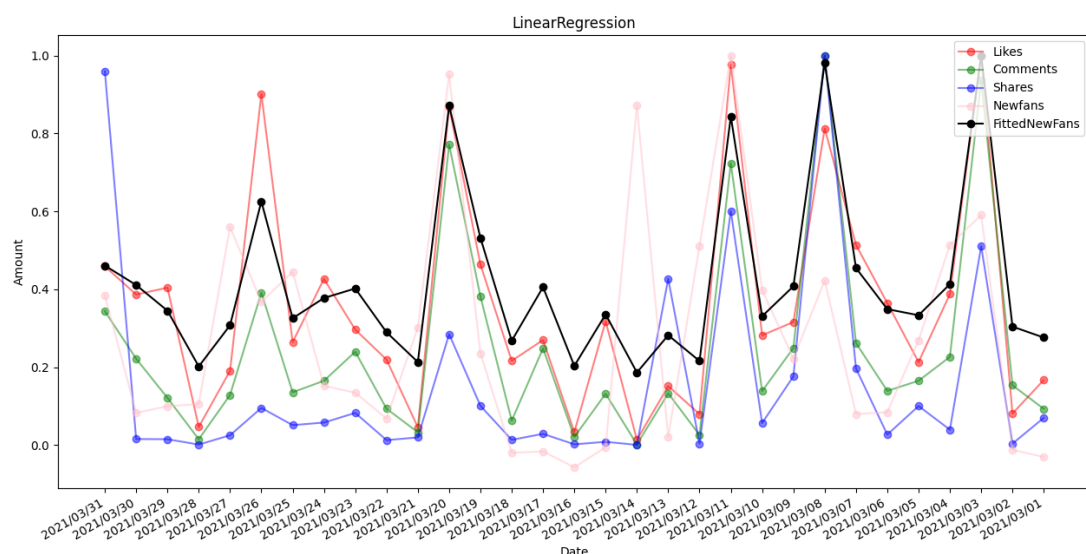
$$\begin{cases} \alpha = 2.48791931e - 03 \\ \beta = 2.17547731e - 01 \\ \gamma = -1.13712150e - 02 \end{cases}$$

其中系数 α 、 β 、 γ 分别对应点赞，评论，转发。

R 方为 0.261 说明拟合效果较好。

作图见 Figure 3

注：由于粉丝增长数、点赞数等数值相差较大，为了能直观显示，我们对数据采取归一化处理。



结论：从系数中可以看出**评论**对关注影响最大，其次是转发，最后是点赞。评论需要打字和思考，耗时更长，对用户来说操作更耗时，如果评论了就说明有较大好感度，也就更愿意点击关注。点赞耗时最少，操作简单，很很可能是不经意间点赞，吸引力也就比较小。这些都符合常理。

文件说明



FanAmountGrap

h.py

粉丝数量绘图



FanGrowthGraph

.py

粉丝增长数量绘图



LinearRegression

.py

多元线性回归分析



NewLinearRegre

ssion.py

归一化多元线性回归分析及绘图



ProcessedData.c

sv

每天的点赞、评论、转发数量